THE PRINCETON ENCYCLOPEDIA OF THE

World Economy

Kenneth A. Reinert
Ramkishen S. Rajan

EDITORS

Amy Jocelyn Glass
Lewis S. Davis

ASSOCIATE EDITORS

The Princeton Encyclopedia of the World Economy

The Princeton

Encyclopedia of the World Economy

EDITORS IN CHIEF

Kenneth A. Reinert

Ramkishen S. Rajan

ASSOCIATE EDITORS

Amy Jocelyn Glass

Lewis S. Davis

PRINCETON UNIVERSITY PRESS

PRINCETON AND OXFORD

The findings, interpretations, and conclusions expressed in this book
are entirely those of the authors writing in their personal capacities.
They do not necessarily represent the views or positions of the in-
stitutions with which the authors are affiliated, or of any other
person associated with those institutions.

This book has been composed in Adobe Garamond and Myriad

Printed on acid-free paper.

Contents

Introduction

The world economy, that collection of human activities spanning national borders, touches us all. The banking executive organizing global financial services and the poor agricultural worker cultivating an export crop are both strongly influenced by global economic forces beyond their control. The banking executive operates within a financial services protocol of the World Trade Organization and in financial markets that increase in importance with every passing year and yet remain stubbornly volatile in new and not always predictable ways. The agricultural worker sees her fortunes rise and fall with the world price of the export crop she produces and the remittances she receives from her son who works abroad in a country she will never visit. The value of these remittances changes with her country's exchange rate, as well as with the business cycle where her son works. The banking executive and the agricultural worker do not know one another, but both are participants in what we call the world economy.

This work conceives of the world economy as the interaction among countries in four broad areas: international trade, international finance, international production, and international economic development. Some decades ago, considerations of the world economy neatly fit into only two of these four areas, namely international trade and international finance. Not today. Significant changes in the way production is taking place in the world, the emergence of new regions as active locations of this production, and conceptual and theoretical advancements in our understanding of economic growth

necessitate a more inclusive view. What we conceive of as the "world economy" here is more than just traditional international economics. Integration among countries in the four broad areas considered here has progressed to the extent that it is appropriate to conceive of the world as having characteristics of a single global economy.

Despite the importance of the modern world economy, people around the globe face a number of challenges in their attempts to understand it. Foremost among these challenges is a tendency for the relevant published works to be either too general or too specialized to suit the needs of students in economics and the broader social sciences, public policy and public affairs, international studies, and business, as well as the large and growing number of professionals working in international economic policy. Works intended for a popular audience tend to focus on broad themes and theories, often with a political point of view. Articles in academic journals often use terminology and mathematics beyond the reach of the nonspecialist.

This encyclopedia is intended to bridge the gap between the general and the specialized. It explains the structure and workings of the world economy, summarizes issues and debates related to economic globalization, and provides suggestions for further reading on more than 300 topics. With a few necessary exceptions, the entries in this book eschew the broad sweep. Instead, we have chosen topics we consider essential to a real understanding of forces at play. The entries for the most part do not employ

advanced mathematics or highly technical language. Their aim is to accurately educate and inform, not overwhelm.

Structure and Coverage

The entries fall into six major categories: concepts and principles; models and theory; institutions and agreements; policies and instruments; analysis and tools; and sectors and special issues. Our selection of topics in the last of these categories, sectors and special issues, was based on policy relevance rather than on theoretical considerations. We have not included biographical entries, but individuals who have played a significant role in the world economy are mentioned in the relevant topical entries; their names also appear in the index.

The *Encyclopedia* is informed by the modern history of international economic relations since the introduction of the gold standard in the late 19th century. This is widely considered to be the beginning of the modern era of economic relations and, as such, offers a useful starting point. We have tried to select topics that will continue to be important   essential concepts, perennial issues, and long-term trends   regardless of the particular future course of the world economy.

International Trade

Around 1980, our understanding of international trade   the exchange of merchandise and services among the countries of the world   began to change. In response to emerging patterns of trade within (rather than between) manufacturing and service sectors, new theories emerged based on imperfect competition and economies of scale, supplementing the old stories of comparative advantage based on factor or resource endowments in which, for example, a country with a relatively large amount of labor would export labor-intensive goods. At the same time, trade policy agendas rapidly expanded into new areas such as trade in services, intellectual property, a new generation of preferential trade agreements, and the settlement of disputes. Trade economists and trade lawyers became acquainted. Unforeseen issues emerged out of or alongside of trade negotiations such as trade and the environment, trade and labor, and trade and public health.

Trade-related entries in this *Encyclopedia* reflect this new reality. Standard models of international trade (the Ricardian model, the Heckscher-Ohlin model, and the specific-factors model) are given their due by world-renowned trade theorists. New Trade Theory (i.e., based on oligopoly and monopolistic competition) also receives attention, as do its applications in areas such as the New Economic Geography. We supplement these core models with entries on a large set of basic concepts, from absolute and comparative advantage to terms of trade and fragmentation. A host of trade policy instruments are covered, from basic tariffs to nontariff measures, including quotas, tariff rate quotas, and technical barriers to trade. A large number of institutions and agreements are covered, from the obvious (e.g., the World Trade Organization) to the less well known (e.g., the Convention on Biological Diversity and the Convention on International Trade in Endangered Species). We also give attention to commonly used tools of analysis, such as revealed comparative advantage, effective protection, and gravity models. Finally, we cover a range of special issues such as access to medicines, gender, and the illegal drugs trade.

International Finance

Changes have been even more dramatic in the realm of international finance than in international trade. The liberalization and integration of global financial markets began in the 1980s, but accelerated significantly in the 1990s. For example, data from the Bank for International Settlements indicate that global foreign exchange turnover increased from U.S. $620 billion in 1990 to U.S. $3.2 trillion in 2007. Potential benefits that can emerge from these changes include improved resource allocation from countries specializing in financial services; increased portfolio

diversification; improved competition in the financial sector; and increased market discipline on policymakers. Such changes can have positive effects on the overall growth and development of the countries involved.

That said, these changes in the landscape of global finance have also been associated with repeated episodes of significant turbulence. For example, in 1992 93, Europe was faced with the very real possibility of a complete collapse of the European Exchange Rate Mechanism (ERM). In 1994 95, the Mexican currency crisis involved a steep devaluation of the peso and brought Mexico to the brink of default, with spillover effects on Argentina and Brazil. Between July 1997 and mid-1998, the world experienced the effects of the East Asian crisis, which started somewhat innocuously with a run on the Thai baht, but spread swiftly to a number of other regional currencies, most notably the Indonesian rupiah, Malaysian ringgit, Philippine peso, and Korean won. Other large emerging economies, such as Russia and Brazil, also experienced periods of significant market weakness and required the assistance of the International Monetary Fund. The Russian ruble was devalued in August 1998, while the Brazilian real's fixed rate to the U.S. dollar was eventually broken in January 1999. A number of other smaller emerging economies, such as Turkey and Ecuador, also experienced currency and financial crises in the 1990s.

Another striking change has been the reversal of capital flows from developed to developing countries. Due in large part to the emergence of a significant current account deficit (i.e. spending in excess of national saving) in the United States and involving the official transactions of central banks, the developing world is now an *exporter* of capital to the developed world rather than an importer. In fact, the flow of international capital from developing countries to developed countries is now one of the key paradoxes of the global economy, as is the fact that foreign governments and central banks have become major participants in global financial markets via the creation of sovereign wealth funds.

Finance-related entries in the *Encyclopedia* reflect these important changes. Standard models of inter-

national finance and open-economy macroeconomics are covered, including the interest parity conditions and the Mundell-Fleming model, and a host of basic concepts such as balance of payments, capital flight, currency crisis, and sterilization. These are supplemented with more recent theoretical contributions such as the New Open Economy Macroeconomics. Entries on policy instruments include capital controls, hedging, and foreign exchange intervention, to name a few. Coverage also includes the basic analytical tools of the field, such as early warning systems and exchange rate forecasting, as well as special topics such as financial services, sequencing of financial sector reform, recycling of petrodollars, and money laundering.

International Production

As mentioned earlier, a key change in the world economy has been in the structures of international production of goods and services, as multinational enterprises (MNEs) engage in varieties of foreign direct investment (FDI) with managerial influence over foreign-based productive enterprises. Drawing on both the work of earlier theory in international business and new developments in trade theory, economists have incorporated MNEs and FDI into the theory and practice of international economics. At the same time, issues such as outsourcing and offshoring have attracted public attention and increased political interest in this area. This has occurred both in high-income countries, where outsourcing has moved to white-collar as well as blue-collar occupation categories, and in developing countries, where new income-generating possibilities have emerged outside of manufacturing.

When deciding to become a MNE, a firm chooses the mode that maximizes its profits, typically considering such options as exporting and licensing a local firm (outsourcing/offshoring) as alternatives. Consequently, models usually revolve around the trade-offs between FDI and at least one alternative. In these models, the location of production differs between FDI and exporting, and whether transactions

occur within the firm differs between FDI and licensing.

Like trade, much FDI occurs between similar, developed countries and is often two-way, but FDI has grown even more rapidly than trade. A large share of world trade (about 30 percent) occurs *within* firms and is known as intrafirm trade. MNEs tend to arise in industries with large research and development expenditures relative to sales, significant product differentiation, and substantial intangible assets such as intellectual property and brand value. FDI is mostly horizontal   with MNEs creating local production facilities in each country and selling within each country or region   rather than vertical, in which MNEs allocate production processes across countries and ship the products back home.

Our coverage of international production begins with issues of theory, from long-standing pursuits such as location theory to emerging issues of importance such as FDI under monopolistic competition and oligopoly. These more formal considerations are again supplemented with a set of entries on basic concepts, from FDI and MNEs themselves to intangible assets and technology licensing. Coverage of policy instruments includes entries on domestic content requirements, foreign equity restrictions, and trade-related investment measures more broadly. We also cover specific analytic tools such as market size and exchange rates as they relate to FDI. Finally, the book addresses several special issues related to FDI, such as FDI and export performance and FDI and labor markets.

International Economic Development

Due to the huge disparities in standards of living around the globe and the tremendous impact, for good or ill, of international economic transactions in the development process, the ideas and policies that shape economic development are subject to ongoing and highly charged debates. In entries on influential development institutions, such as the World Bank, and policy frameworks, such as import substitution industrialization and the Washington consensus, our

goal is to explain these debates without attempting to resolve them. Entries on economic development and the evolution of development thinking are designed to provide the reader with a conceptual framework and a context for understanding the issues addressed in entries such as international migration, international trade and economic development, and technological progress in open economies.

Since the 1990s, development thinking has shifted toward recognizing the central importance of political processes in structuring and potentially limiting economic development. Reflecting this shift, the *Encyclopedia* attempts to view the economic processes of development in the broader context of political decision-making. This emphasis on political economy is apparent in many of the entries on economic policy and takes center stage in entries that address the linkages between economic and political systems, such as those on democracy and development, corruption, and international aid and political economy. Finally, while most of the development entries take the national economy as their starting point, entries on global poverty, global income inequality, and international income convergence take a deliberately transnational approach, reflecting the growing consensus that it is now meaningful to view some issues from the perspective of a single global economy.

Using the Encyclopedia

There are a number of ways to engage with the encyclopedia. Entries are arranged in alphabetical order. If you are uncertain of the particular entry that may be most appropriate to your interest, you can consult the Topical List of Entries, which groups the entry terms by category. You also can consult the index to find topics that are covered in related entries but do not have entries of their own.

At the end of most entries you will find "See also" references to other entries related to the topic at hand. Each entry concludes with an annotated Further Reading list to guide readers with an interest in additional research.

Acknowledgments

Alphabetical List of Entries

Topical List of Entries

1. Concepts and Principles
2. Models and Theory
3. Institutions and Agreements
4. Policies and Instruments
5. Analysis and Tools
6. Sectors and Special Issues

1. Concepts and Principles

International Trade
absolute advantage
anti-globalization
commodity chains
common market
comparative advantage
competitive advantage
customs union
distortions to agricultural incentives
fair trade
fragmentation
free trade area
gains from trade
globalization
intrafirm trade
intraindustry trade
migration, international
multilateralism
outsourcing/offshoring
parallel imports
regionalism

temporary movement of natural
   persons
terms of trade
trade in services
trade-related capacity building
transfer problem

International Finance
assignment problem
asymmetric information
bail-ins
bailouts
balance of payments
balance sheet approach/effects
banking crisis
black market premium
bubbles
capital controls
capital flight
capital mobility
carry trade
common currency
conflicted virtue
contagion
convertibility
currency competition
currency crisis
debt deflation
deposit insurance
discipline
dominant currency

2. Models and Theory

International Trade
economies of scale
foreign direct investment under monopolistic
    competition
foreign direct investment under oligopoly
Heckscher-Ohlin model
monopolistic competition
New Economic Geography
New Trade Theory
nontraded goods
oligopoly models
political economy of trade policy
Ricardian model
specific-factors model

International Finance
Balassa-Samuelson effect
Feldstein-Horioka puzzle
forward premium puzzle
interest parity conditions
J-curve effect
Marshall-Lerner condition
Mundell-Fleming model
New Open Economy Macroeconomics
optimum currency area (OCA) theory
primary products trade
quantity theory of money
Swan diagram
Triffin dilemma

International Production
foreign direct investment: the OLI framework
foreign direct investment under monopolistic
    competition
foreign direct investment under oligopoly
internalization theory
knowledge-capital model of the multinational
    enterprise
location theory
pollution haven hypothesis
proximity-concentration hypothesis

International Economic Development
aid, international, and political economy
dependency theory
evolution of development thinking
growth in open economies, neoclassical models
growth in open economies, Schumpeterian
    models
infant industry argument
North-South trade
South-South trade

3. Institutions and Agreements

International Trade
African Caribbean Pacific–European Union
    partnership agreements (ACP-EU)
African Union
Agreement on Agriculture
Agreement on Trade-Related Aspects of
    Intellectual Property Rights (TRIPS)
agricultural trade negotiations
Andean Community
Asia Pacific Economic Cooperation (APEC)
Association of Southeast Asian Nations (ASEAN)
Basel Convention
Central American Common Market (CACM)
Central American–Dominican Republic Free
    Trade Area (CAFTA-DR)
Common Agricultural Policy
Common Market for Eastern and Southern Africa
    (COMESA)
Convention on Biological Diversity
Convention on International Trade in
    Endangered Species (CITES)
Doha Round
Economic Community of West African States
    (ECOWAS)
European Union
Free Trade Area of the Americas (FTAA)
General Agreement on Tariffs and Trade (GATT)
General Agreement on Trade in Services (GATS)
Global Environment Facility
Group of Seven/Eight (G7/G8)

Directory of Contributors

Rajat Acharyya
Reader, Department of Economics, Jadavpur University, Calcutta
   trade and wages (coauthor)

Charles Adams
Visiting Professor, Lee Kuan Yew School of Public Policy, National University of Singapore
   carry trade

Joshua Aizenman
Professor of Economics, University of California, Santa Cruz
   financial crisis; international reserves

Mark Allen
Director, Policy and Development Review Department, International Monetary Fund
   balance sheet approach/effects (coauthor)

Kym Anderson
George Gollin Professor of Economics, University of Adelaide
   distortions to agricultural incentives

Klaus Armingeon
Professor of Political Science and Director, Institute of Political Science, University of Berne
   Organisation for Economic Co-operation and Development (OECD)

Sven W. Arndt
Charles M. Stone Professor of Money, Credit, and Trade, Claremont McKenna College
   fragmentation

Nancy Neiman Auerbach
Associate Professor of International Political Economy, Scripps College
   financial liberalization; Latin American debt crisis; petrodollars, recycling of (coauthor)

Marc Bacchetta
Counselor, Economic Research and Statistics Division World Trade Organization
   tariff escalation; tariffs

Christopher M. Bacon
Lecturer, Departments of Latin American and Latino Studies, Sociology and Environmental Studies, University of California, Santa Cruz
   fair trade

Jennifer Bair
Assistant Professor of Sociology, Yale University
   commodity chains

King Banaian
Professor of Economics, St. Cloud State University
   currency substitution and dollarization

Subhayu Bandyopadhyay
Research Officer, Federal Reserve Bank of St. Louis

common market (coauthor)

Frank Barry
*Lecturer, School of Economics,
University College Dublin*

foreign direct investment and labor markets

Andrew L. Beath
*PhD Candidate in Government, Harvard
University*

migration, international (coauthor)

Hamid Beladi
*Professor of Economics, College of Business,
University of Texas at San Antonio*

nontraded goods (coauthor)

Paul Bergin
*Associate Professor of Economics, University
of California, Davis*

Balassa-Samuelson effect

Pradip Bhatnagar
*Economist, Technical Support Facility of the Group
of Fifteen (G-15), Geneva*

temporary movement of natural persons

Richard Blackhurst
*Professor of International Economics,
The Fletcher School, Tufts University*

General Agreement on Tariffs and Trade (GATT);
World Trade Organization

Robert A. Blecker
Professor of Economics, American University

steel

Clemens F. J. Boonekamp
*Director, Trade Policies Review Division,
World Trade Organization*

Trade Policy Review Mechanism

James M. Boughton
Historian, International Monetary Fund

International Monetary Fund surveillance

François Bourguignon
Director, Paris School of Economics

global income inequality (coauthor)

Paul Bowles
*Professor of Economics, University of Northern British
Columbia*

beggar-thy-neighbor policies; mercantilism

Russell S. Boyer
*Professor of Economics, University of Western
Ontario*

Mundell-Fleming model

Carlos A. Primo Braga
*Senior Advisor, International Policy and Partnerships
Group, World Bank*

World Trade Organizaton, accession to
(coauthor)

Mark R. Brawley
Professor of Political Science, McGill University

globalization

Drusilla K. Brown
Associate Professor of Economics, Tufts University

labor standards

Richard C. K. Burdekin
*Jonathan B. Lovelace Professor of Economics,
Claremont McKenna College*

commodity-price pegging; debt deflation;
quantity theory of money; seigniorage

Kenneth Button
*Professor of Public Policy and Director, Center for
Transportation Policy, Operations, and Logistics,
George Mason University*

air transportation; shipping

Fang Cai
Economist, Division of International Finance,
Federal Reserve

  home country bias

Rui Castro
Associate Professor of Economics,
Université de Montréal

  growth in open economies, neoclassical models

Tony Cavoli
Lecturer, School of Commerce, University of
South Australia

  capital mobility (coauthor); inflation targeting;
  monetary policy rules

Avik Chakrabarti
Associate Professor of Economics,
University of Wisconsin–Milwaukee

  nontraded goods (coauthor)

Maggie Xiaoyang Chen
Assistant Professor of Economics and International
Affairs, George Washington University

  fixed costs and foreign direct investment;
  market access

Yin-Wong Cheung
Professor of Economics, University of California,
Santa Cruz

  exchange rate forecasting; purchasing power
  parity

Menzie D. Chinn
Professor of Economics and Public Affairs, Robert M.
La Follette School of Public Affairs, University of
Wisconsin

  effective exchange rate; forward premium
  puzzle; interest parity conditions

Eric M. P. Chiu
Assistant Professor, Institute of National Policy and
Public Affairs, National Chung Hsing University

  currency crisis (coauthor)

Christopher K. Clague
Emeritus Professor of Economics,
University of Maryland

  democracy and development

Jennifer Clapp
CIGI Chair in International Governance and Professor,
Faculty of Environmental Studies,
University of Waterloo

  Basel Convention

Kimberly A. Clausing
Professor of Economics, Reed College

  intrafirm trade

Paul Collier
Professor of Economics and Director, Center for the
Study of African Economies, University of Oxford

  aid, military (coauthor)

Brian R. Copeland
Professor of Economics, University of British
Columbia

  pollution haven hypothesis

Germán Creamer
Center for Computational Learning Systems,
Columbia University

  Andean Community

Edward Crenshaw
Associate Professor of Sociology,
Ohio State University

  digital divide

Michael R. Curtis
Adjunct Professor of Public Policy,
George Mason University

  petroleum

Lewis S. Davis
Assistant Professor of Economics, Union College

  development

Harry de Gorter
Associate Professor, Department of Applied Economics and Management, Cornell University

    tariff rate quotas

James W. Dean
Ross Distinguished Professor, Department of Economics, Western Washington University

    Eurocurrencies; exorbitant privilege; money supply

Alan V. Deardorff
John W. Sweetland Professor of International Economics and Professor of Economics and Public Policy, University of Michigan

    Ricardian model

Gelaye Debebe
Assistant Professor of Organizational Sciences, George Washington University

    brain waste

Sirathorn Dechsakulthorn
PhD Candidate in Economics, Claremont Graduate University

    discipline (coauthor)

Robert Dekle
Professor of Economics, University of Southern California

    banking crisis

Desmond Dinan
Jean Monnet Professor, School of Public Policy, George Mason University

    European Union

Elias Dinopoulos
Professor of Economics, University of Florida

    growth in open economies, Schumpeterian models

Jonathan P. Doh
Associate Professor of Management and Operations, Herbert G. Rammrath Endowed Chair in International Business, and Director, Center for Global Leadership, Villanova School of Business, Villanova University

    nongovernmental organizations (NGOs)

Kathryn M. E. Dominguez
Professor of Public Policy and Economics, University of Michigan, and Research Associate, National Bureau of Economic Research

    sterilization

Brian Doyle
Senior Economist, International Finance Division, Federal Reserve Board

    New Open Economy Macroeconomics

Graham Dutfield
Professor of International Governance, School of Law, University of Leeds

    Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS); World Intellectual Property Organization

Eric V. Edmonds
Associate Professor of Economics, Dartmouth College; Director, Child Labor Network, Institute for the Study of Labor (IZA); Faculty Research Fellow, National Bureau of Economic Research

    child labor

Andrew Elek
Independent Scholar

    Asia Pacific Economic Cooperation (APEC) (coauthor)

Kimberly Ann Elliott
Senior Fellow, Center for Global Development and Peter G. Peterson Institute for International Economics

    International Labor Organization

Korkut A. Erturk
Professor of Economics, University of Utah
bubbles; speculation

Wilfred J. Ethier
Professor of Economics, University of Pennsylvania
economies of scale

Simon J. Evenett
Professor of International Trade and Economic Development, Swiss Institute for International Economics and Applied Economic Research, University of St. Gallen; Research Fellow, Center for Economic Policy Research (CEPR)
competition policy; government procurement; World Trade Organization, accession to (coauthor)

David Favre
Professor of Law, Michigan State University College of Law
Convention on International Trade In Endangered Species (CITES)

Robert C. Feenstra
C. Bryan Cameron Distinguished Chair in International Economics and Director, Center for International Data, University of California, Davis; Director, International Trade and Investment Program, National Bureau of Economic Research
outsourcing/offshoring (coauthor)

Richard Feinberg
Professor of International Political Economy and Director, APEC Study Center, University of California, San Diego
Free Trade Area of the Americas (FTAA)

Susan Feinberg
Associate Professor of International Business, Rutgers Business School
location theory

Michael Ferrantino
International Economist, U.S. International Trade Commission
appropriate technology and foreign direct investment

Francisco Ferreira
Development Research Group, World Bank
global income inequality (coauthor)

Ben Ferrett
Research Fellow, Leverhulme Center for Research on Globalization and Economic Policy, School of Economics, University of Nottingham
foreign direct investment under oligopoly

Christopher Findlay
Professor of Economics, University of Adelaide
Asia Pacific Economic Cooperation (APEC) (coauthor)

Lionel Fontagné
Professor of Economics, Paris School of Economics, Université de Paris 1 Panthéon-Sorbonne
South-South trade

Joseph F. Francois
Professor of Economics, Johannes Kepler Universität Linz
partial equilibrium models (coauthor)

Michele Fratianni
Professor of Business Economics and Public Policy Emeritus, Indiana University, and Professor of Economics, Università Politecnica delle Marche, Ancona
Bank for International Settlements (BIS); dominant currency; Feldstein-Horioka puzzle; gold standard, international

Francis Fukuyama
Bernard L. Schwartz Professor of International Political Economy and Director, International

Development Program, Paul H. Nitze School of Advanced International Studies, Johns Hopkins University

    international institutional transfer

Andrea K. Gerlak

Visiting Professor, Department of Political Science, University of Arizona

    Global Environment Facility (coauthor)

Chetan Ghate

Assistant Professor of Economics, Indian Statistical Institute, New Delhi

    time inconsistency problem

Amit Ghosh

Assistant Professor, Department of Economics, Illinois Wesleyan University

    exchange rate pass-through

Christopher L. Gilbert

Professor of Econometrics, University of Trento

    World Bank (coauthor)

Kate Gillespie

Research Associate Professor, Department of Marketing, McCombs School of Business, University of Texas at Austin

    smuggling

Amy Jocelyn Glass

Associate Professor of Economics, Texas A&M University

    foreign direct investment and innovation, imitation (coauthor); infrastructure and foreign direct investment; vertical versus horizontal foreign direct investment

Linda S. Goldberg

Federal Reserve Bank of New York and National Bureau of Economic Research

    exchange rates and foreign direct investment

Ian Goldin

Director, James Martin 21st Century School, and Professorial Fellow, Balliol College, University of Oxford

    aid, international; evolution of development thinking; migration, international (coauthor)

Holger Görg

Associate Professor and Reader, School of Economics, University of Nottingham

    foreign direct investment and exit of local firms; linkages, backward and forward

Ashima Goyal

Professor of Economics, Indira Gandhi Institute of Development Research, Mumbai, India

    assignment problem; hedging; hot money and sudden stops (coauthor); Swan diagram

Ilene Grabel

Professor of International Economics, Graduate School of International Studies, University of Denver

    Tobin tax

Jean-Christophe Graz

Assistant Professor, Institute of Political and International Studies, University of Lausanne

    World Economic Forum

Duncan Green

Head of Research for Oxfam GB

    anti-globalization (coauthor)

Joshua Greene

Deputy Director, IMF-Singapore Regional Training Institute

    black market premium; dual exchange rate

Matt Griffith

Senior Advisor, Commission for Rural Communities in London

    anti-globalization (coauthor)

Stephany Griffith-Jones
Executive Director of the Initiative for Policy Dialogue, Columbia University

   capital flows to developing countries

Nigel Grimwade
Principal Lecturer of Economics and Finance, Department of Business, Computing, and Information Management, London South Bank University

   competitive advantage

Alfred V. Guender
Associate Professor, Department of Economics, University of Canterbury

   monetary conditions index

Morley Gunderson
Professor, Center for Industrial Relations and Human Resources and Department of Economics, University of Toronto; CIBC Chair in Youth Employment

   social policy in open economies

Markus Haacker
Economist, African Department, International Monetary Fund

   health and globalization; HIV/AIDS; Millennium Development Goals

David M. Hart
Associate Professor, School of Public Policy, George Mason University

   brain gain

C. Randall Henning
Professor, School of International Service at American University and Visiting Fellow, Peter G. Peterson Institute for International Economics

   exchange rate weapon

Almas Heshmati
Professor of Economics and Head of Department of Economics and Finance, University of Kurdistan

Hawler; Research Fellow, Techno-Economics, Management and Policy Program, Seoul National University

   information and communications technology (coauthor)

Jack High
Professor, School of Public Policy, George Mason University

   balance of payments

Anke Hoeffler
Research Officer, Center for the Study of African Economics, Department of Economics, St. Antony's College, University of Oxford

   aid, military (coauthor)

Bernard Hoekman
Senior Advisor, Development Research Group, World Bank and Researcher, Center for Economic Policy Research (CEPR)

   Doha Round; Uruguay Round

Henrik Horn
Senior Research Fellow, Research Institute of Industrial Economics (IFN), Stockholm, and Research Fellow, Center for Economic Policy Research (CEPR)

   nondiscrimination (coauthor)

Brett House
Economist, Policy and Development Review Department, International Monetary Fund

   balance sheet approach/effects (coauthor)

Andrew Hughes Hallett
Professor, School of Public Policy, George Mason University

   Bonn Summit; international policy coordination; Smithsonian Agreement

Michael Hutchison
Professor of Economics, University of California, Santa Cruz

   Bank of Japan; exchange rate regimes

Barry W. Ickes
Professor of Economics, Pennsylvania State University

   political economy of policy reform

Asif Islam
Research Associate, Department of Agricultural and Resource Economics, University of Maryland

   trade and the environment (coauthor)

Hiro Ito
Associate Professor, Department of Economics, Portland State University

   expenditure changing and expenditure switching; financial repression; liquidity trap, the

J. Bradford Jensen
Associate Professor of Economics and International Business, McDonough School of Business, Georgetown University; Senior Fellow at Peter G. Peterson Institute for International Economics; Research Associate at National Bureau of Economic Research

   outsourcing/offshoring (coauthor)

Omotunde E. G. Johnson
Retiree, International Monetary Fund, and Adjunct Professor, School of Public Policy, George Mason University

   African Union

R. Barry Johnston
Assistant Director, Monetary and Capital Markets Department, International Monetary Fund

   international financial centers; money laundering; offshore financial centers; sequencing of financial sector reforms

Ronald W. Jones
Xerox Professor of Economics, University of Rochester

   specific-factors model

Philippe Jorion
Chancellor's Professor of Finance, The Paul Merage School of Business, University of California at Irvine

   hedge funds

Tim Josling
Senior Fellow, Freeman Spogli Institute for International Studies, Stanford University, and Visiting Professor, Imperial College, London

   Agreement on Agriculture; agricultural trade negotiations

Joseph P. Joyce
Professor of Economics, Wellesley College

   International Monetary Fund (IMF); International Monetary Fund conditionality

William H. Kaempfer
Vice Provost and Professor of Economics, University of Colorado, Boulder

   sanctions (coauthor)

Inge Kaul
Special Advisor, United Nations Development Program

   global public goods

Sydney J. Key
Lecturer, Graduate Program in Banking and Financial Law, Boston University School of Law

   financial services

Christopher Kilby
Associate Professor of Economics, Vassar College

   aid, bilateral

John Kirton
Associate Professor of Political Science, University of Toronto

   Group Of Seven/Eight (G7/G8)

Jörn Kleinert
Professor of Economics, Eberhard-Karls-University Tübingen

   market size and foreign direct investment

Vladimir Klyuev
Economist, Western Hemisphere Department, International Monetary Fund

   capital accumulation in open economies

Ulrich Koester
Professor of Agricultural Economics, University of Kiel, Germany

    Common Agricultural Policy

Michel Kostecki
Professor of Marketing, Enterprise Institute, University of Neuchâtel

    trade-related capacity building

Jan Kregel
Distinguished Research Professor in the Center for Full Employment and Price Stability, University of Missouri, Kansas City; Senior Scholar, Levy Economics Institute, Bard College

    Bretton Woods system; Triffin dilemma

Kala Krishna
Professor of Economics, Pennsylvania State University

    free trade area; rules of origin

Pravin Krishna
Chung Ju Yung Distinguished Professor of International Economics and Business, Paul H. Nitze School of Advanced International Studies, and Professor of Economics, Johns Hopkins University

    multilateralism; regionalism

Maurice Kugler
Center for International Governance Innovation Chair in International Public Policy and Professor of Economics, Wilfrid Laurier University, and Visiting Professor of Public Policy, John F. Kennedy School of Government, Harvard University

    factor endowments and foreign direct investment

Philip R. Lane
Professor of Economics, and Director, Institute for International Integration Studies, Trinity College, Dublin

    transfer problem

Minkyu Lee
PhD Candidate, Technology Management Economics and Policy Program, Seoul National University

    information and communications technology (coauthor)

James Lehman
Professor of Economics, Pitzer College

    Federal Reserve Board

Priscilla Liang
Assistant Professor of Business and Economics, California State University, Channel Islands

    contagion (coauthor)

Ramón López
Professor of Economics, Department of Agriculture and Resource Economics, University of Maryland

    trade and the environment (coauthor)

Eduardo Lora
Principal Advisor, Research Department, Inter-American Development Bank

    Washington consensus

Mary E. Lovely
Associate Professor of Economics, Syracuse University

    agglomeration and foreign direct investment

Anton D. Lowenberg
Professor of Economics, California State University, Northridge

    sanctions (coauthor)

Rodney Ludema
Associate Professor of Economics, Georgetown University

    multilateral trade negotiations

Ronald MacDonald
Adam Smith Chair of Political Economy and Director of PhD Programs, Department of Economics, University of Glasgow

equilibrium exchange rate; exchange rate volatility

Alex Mandilaras
Lecturer in Economics, University of Surrey

early warning systems; exchange market pressure

Andrea Maneschi
Professor of Economics and Director, Graduate Program in Economic Development, Vanderbilt University

comparative advantage; gains from trade

Sugata Marjit
Director and Reserve Bank of India Professor, Center for Studies in Social Sciences, Calcutta

trade and wages (coauthor)

Stephen Marks
Professor of Economics, Pomona College

asymmetric information

James R. Markusen
Professor of International Economics, University of Colorado

foreign direct investment (FDI); internalization theory

Antoine Martin
Senior Economist, Federal Reserve Bank of New York

currency competition; multiple currencies

Philip Martin
Professor of Agricultural and Resource Economics and Chair, UC Comparative Immigration and Integration Program, University of California, Davis

migration governance

Will Martin
Lead Economist, Trade and Development Research Group, World Bank

tariff-cutting formulas

Catherine Matraves
Visiting Assistant Professor of Economics, Michigan State University

pharmaceuticals

Xenia Matschke
Assistant Professor of Economics, University of Connecticut

New Trade Theory

Aaditya Mattoo
Lead Economist, Development Research Group, World Bank

General Agreement on Trade in Services (GATS); trade in services

Petros C. Mavroidis
Edwin B. Barker Professor of Law, Columbia Law School; Professor of Law, University of Neuchâtel

nondiscrimination (coauthor)

Daniel Maxwell
Associate Professor, The Fletcher School, Tufts University

aid, food; aid, humanitarian

Joseph A. McCahery
Professor of Corporate Governance, Center for Law and Economics, University of Amsterdam

corporate governance

Ronald I. McKinnon
William D. Eberle Professor of International Economics, Stanford University

conflicted virtue; dollar standard

Niall P. Meagher
Senior Counsel, Advisory Center on World Trade Organization Law, Geneva

World Trade Organization dispute settlement (coauthor)

Jose Mendez
Professor of Economics, W. P. Carey School of Business, Arizona State University

    Central American Common Market (CACM); Central American–Domincan Republic Free Trade Area (CAFTA-DR)

Mokhtar M. Metwally
Professor of Economics, University of Wollongong

    Gulf Cooperation Council

Klaus Meyer
Professor of Strategy and International Business, School of Management, University of Bath

    foreign market entry

Constantine Michalopoulos
Independent Consultant

    special and differential treatment

Branko Milanovic
Economist, Development Research Group, World Bank

    global income inequality (coauthor)

Chris Milner
Assistant Professor, School of Economics and Leverhulme Center for Research in Globalization and Economic Policy (GEP), University of Nottingham

    effective protection; export promotion

Eugenio J. Miravete
Associate Professor of Economics, University of Texas

    infant industry argument

Ajit Mishra
Senior Lecturer, Department of Economics and International Development, University of Bath

    corruption

Andrew Mitchell
Senior Lecturer, Melbourne Law School

    electronic commerce

Devashish Mitra
Professor of Economics and Gerald B. and Daphna Cramer Professor of Global Affairs, Maxwell School, Syracuse University

    trade and economic development, international

Pradeep K. Mitra
Chief Economist, Europe and Central Asia Region, World Bank

    transition economies

José G. Montalvo
Professor, Department of Economics, Universitat Pompeu Fabra

    aid, international, and political economy

Michael O. Moore
Professor of Economics and International Affairs and Director, Institute for Economic Policy, Elliott School of International Affairs, George Washington University

    anti-dumping

C. W. Morgan
Associate Professor and Research Fellow, Center for Economic Development and International Trade (CREDIT), School of Economics, University of Nottingham

    African Caribbean Pacific–European Union (ACP-EU) partnership agreements

Arijit Mukherjee
Associate Professor and Reader in Economics, University of Nottingham

    technology licensing

Jacob Wanjala Musila
Associate Professor of Economics, Athabasca University

    Common Market for Eastern and Southern Africa (COMESA); Economic Community of West African States (ECOWAS)

xxxv

Ethan C. Myers
Teaching Associate, Department of Social Thought and Political Economy, University of Massachusetts

    Global Environment Facility (coauthor)

Usha Nair-Reichert
Associate Professor of Economics and Undergraduate Director, School of Economics, Georgia Institute of Technology

    proximity-concentration hypothesis

Rajneesh Narula
Professor of International Business Regulation, Department of Economics, University of Reading

    multinational enterprises

Léonce Ndikumana
Associate Professor of Economics, University of Massachusetts at Amherst

    capital flight

J. Peter Neary
Professor of Economics, University of Oxford and Research Fellow, Center for Economic Policy Research (CEPR)

    foreign direct investment: the OLI framework

Douglas Nelson
Professor of Economics, Tulane University

    political economy of trade policy

Richard Newfarmer
Economic Advisor, International Trade and Department and Prospects Group, World Bank

    international investment agreements

Hildegunn Kyvik Nordås
Senior Economist, Organisation for Economic Co-operation and Development (OECD)

    textiles and clothing

Ilan Noy
Associate Professor of Economics, University of Hawai'i, Manoa

capital controls (coauthor); hot money and sudden stops (coauthor); special drawing rights

Gianmarco I. P. Ottaviano
Professor of Economics, University of Bologna

    New Economic Geography

David Palmeter
Senior Counsel, Sidley Austin LLP, Washington, DC

    World Trade Organization dispute settlement (coauthor)

Alberto Paloni
Senior Lecturer, Department of Economics, and Director, Center for Development Studies, Glasgow University

    structural adjustment

Arvind Panagariya
Professor of Economics and Jagdish Bhagwati Professor of Indian Political Economy, School of International and Public Affairs, Columbia University

    Heckscher-Ohlin model

Letizia Paoli
Professor, Leuven Institute of Criminology, K.U. Leuven Faculty of Law, Belgium

    illegal drugs trade

Giovanni Peri
Associate Professor of Economics, University of California, Davis

    techonology spillovers

Emile-Robert Perrin
High Council on International Co-operation (HCCI), French Prime Minister's Office

    regional development banks (coauthor)

Richard Pomfret
Professor of Economics, University of Adelaide

    common currency; optimum currency area (OCA) theory

Helen Popper
Associate Professor of Economics, Santa Clara University

  foreign exchange intervention; real exchange rate

Susan Pozo
Professor of Economics, Western Michigan University

  remittances

Thitima Puttitanun
Assistant Professor of Economics, San Diego State University

  intellectual property rights and foreign direct investment

Horst Raff
Professor of Economics, Christian-Albrechts University of Kiel

  joint ventures; mergers and acquisitions

Ramkishen S. Rajan
Associate Professor, School of Public Policy, George Mason University

  capital controls (coauthor); capital mobility (coauthor); exchange rate pass-through (coauthor); fear of floating (coauthor); hot money and sudden stops (coauthor); sovereign wealth funds (coauthor)

Carlos D. Ramirez
Associate Professor of Economics, George Mason University

  convertibility; original sin

Priya Ranjan
Associate Professor of Economics, University of California, Irvine

  trade costs and foreign direct investment

Martin Ravallion
Director, Development Research Group, World Bank

  global income inequality (coauthor)

Kenneth A. Reinert
Professor, School of Public Policy, George Mason University

  applied general equilibrium models; gravity models; migration, international (coauthor); partial equilibrium models (coauthor)

Nola Reinhardt
Professor of Economics, Smith College

  primary products trade

Kara M. Reynolds
Assistant Professor of Economics, American University

  countervailing duties

Francisco Rivera-Batiz
Professor of Economics and Education, Teachers College, Columbia University

  economic development

Donna Roberts
Senior Economist, Economic Research Service, U.S. Department of Agriculture

  sanitary and phytosanitary measures

G. Chris Rodrigo
Independent Consultant

  revealed comparative advantage

Riordan Roett
Sarita and Don Johnston Professor of Political Science and Director of Western Hemisphere Studies, Paul H. Nitze School of Advanced International Studies (SAIS), Johns Hopkins University

  Mercosur

Christoph Rosenberg
Senior Regional Representative for Central Europe and the Baltics, European Department, International Monetary Fund

  balance sheet approach/effects (coauthor)

G. Kristin Rosendal
Senior Research Fellow, Fridtjof Nansen Institute
  Convention on Biological Diversity

Roy J. Ruffin
Anderson Professor of Economics,
University of Houston
  monopolistic competition; oligopoly

C. Ford Runge
Distinguished McKnight University Professor of
Applied Economics and Law and Director, Center for
International Food and Agricultural Policy, University
of Minnesota
  multilateral environmental agreements

Kamal Saggi
Dedman Distinguished Collegiate Professor
of Economics, Southern Methodist University
  foreign direct investment and innovation,
  imitation (coauthor); foreign direct investment
  and international technology transfer

Dominick Salvatore
Distinguished Professor and Director of the PhD
Program, Department of Economics, Fordham
University
  euro; global imbalances; international liquidity;
  reserve currency; vehicle currency

Carlos Santiso
Governance and Public Finance Adviser, Department
for International Development (DFID),
United Kingdom
  regional development banks (coauthor)

John Sargent
Assistant Professor of Management, College of
Business, University of Texas-Pan American
  export processing zones

Prabirjit Sarkar
Professor of Economics, Jadavpur University,
Calcutta
  terms of trade

Frederic M. Scherer
Aetna Professor of Public Policy Emeritus,
John F. Kennedy School of Government, Harvard
University
  access to medicines

Nicolas Schmitt
Professor of Economics, Simon Fraser University;
Research Fellow, Center for Economic Studies, Ifo
Institute for Economic Research, Munich
  parallel imports

Patricia Higino Schneider
Assistant Professor of Economics, Mount Holyoke
College
  intellectual property rights

Jeffrey J. Schott
Senior Fellow, Peter G. Peterson Institute for
International Economics
  North American Free Trade Agreement (NAFTA)
  (coauthor)

M. Fuat Sener
Associate Professor of Economics, Union College
  technological progress in open economies

Brad Setser
Fellow for Geoeconomics, Council on Foreign
Relations
  bail-ins; bailouts; international financial
  architecture

Pierre L. Siklos
Professor of Economics and Director, Viessmann
European Research Center, Wilfrid Laurier
University
  European Central Bank; European Monetary
  Union; impossible trinity; Maastricht Treaty

Keith Sill
Senior Economist, Federal Reserve Bank
of Philadelphia
  peso problem; sovereign risk

Slavi Slavov
Assistant Professor of Economics,
Pomona College

   currency board arrangement (CBA); fear of floating (coauthor)

James L. Smith
Cary M. Maguire Chair in Oil and Gas
Management, Department of Finance,
Edwin L. Cox School of Business,
Southern Methodist University

   Organization of the Petroleum Exporting Countries (OPEC)

Pamela J. Smith
Associate Professor, Department of Applied
Economics, University of Minnesota

   trade-related investment measures (TRIMs)

Servaas Storm
Faculty of Technology, Policy and Management,
Department of Economics, Delft University of
Technology

   import substitution industrialization

Deborah L. Swenson
Associate Professor of Economics, University of
California, Davis

   domestic content requirements; transfer pricing

Alan O. Sykes
Professor of Law, Stanford Law School

   safeguards

Wendy Takacs
Professor of Economics, University of Maryland,
Baltimore County

   nontariff measures; quotas

Evan Tanner
Senior Economist, Western Hemisphere Division,
International Monetary Fund

   monetary versus fiscal dominance

Vito Tanzi
Former Director of the Fiscal Affairs Department,
International Monetary Fund

   subsidies and financial incentives to foreign direct investment

David G. Tarr
Consultant and former Lead Economist,
Development Research Group, World Bank

   customs unions

Jonathan R. W. Temple
Professor of Economics, University of Bristol

   international income convergence

Willem Thorbecke
Associate Professor of Economics, George Mason
University; Senior Fellow, Research Institute
of Economy, Trade, and Industry, Tokyo

   Louvre Accord; Plaza Accord; twin deficits

Jose L. Tongzon
Associate Professor of Economics, National University
of Singapore

   Association of Southeast Asian Nations (ASEAN)

Farid Toubal
Assistant Professor, Paris School of Economics,
University of Paris I, Panthéon-Sorbonne, CNRS,
France

   intangible assets

John Toye
Visiting Professor of Economics, St. Antony's College,
University of Oxford

   United Nations Conference on Trade and Development

Dieter M. Urban
Professor, Department of Business and Economics,
Johannes Gutenberg-Universität

   knowledge-capital model of the multinational enterprise

Diego E. Vacaflores
Assistant Professor, Department of Finance and Economics, Texas State University

    foreign direct investment and tax revenues

Hylke Vandenbussche
Professor of International Economics, Université Catholique de Louvain

    footloose production

Charles van Marrewijk
Professor of Economics, Erasmus University Rotterdam

    absolute advantage; intraindustry trade

Irene van Staveren
Associate Professor of Feminist Development Economics, Institute of Social Studies, The Hague; Professor of Economics and Christian Ethics, Institute for Management Research, Radboud University Nijmegen

    gender

Anthony J. Venables
Professor of Economics, University of Oxford

    foreign direct investment under monopolistic competition

Matias Vernengo
Assistant Professor of Economics, University of Utah

    dependency theory

David Vines
Professor of Economics, Oxford University

    World Bank (coauthor)

Peter Walkenhorst
Senior Economist, International Trade Group, World Bank

    foreign equity restrictions

Howard J. Wall
Vice President and Regional Economics Advisor, Federal Reserve Bank of St. Louis

    common market (coauthor)

Stefanie Walter
Researcher, Center for Comparative and International Studies, Swiss Federal Institute of Technology, Zurich

    currency crisis (coauthor)

Joshua C. Walton
PhD Candidate, School of Politics and Economics, Claremont Graduate University

    currency crisis (coauthor)

Howard White
Fellow, Institute of Development Studies, University of Sussex

    poverty, global

Clas Wihlborg
Fletcher Jones Chair in International Business, Argyros School of Business and Economics, Chapman University

    deposit insurance; lender of last resort

Thomas D. Willett
Director of the Claremont Institute for Economic Studies and Horton Professor of Economics, Claremont Graduate University and Claremont McKenna College

    contagion (coauthor); currency crisis (coauthor); discipline (coauthor); petrodollars, recycling of (coauthor); sovereign wealth funds (coauthor)

John Williamson
Senior Fellow, Peter G. Peterson Institute for International Economics

    band, basket, and crawl (BBC)

John S. Wilson
Lead Economist, Development Research Group, International Trade, World Bank

    technical barriers to trade; trade facilitation

Peter Wilson
Independent Scholar

    J-curve effect; Marshall-Lerner condition

L. Alan Winters

Professor of Economics, University of Sussex; Research Fellow and former Program Director of the Center for Economic Policy Research (CEPR)

    North-South trade

Paul Winters

Associate Professor of Economics, American University

    agriculture

Kar-yiu Wong

Professor of Economics and Director, Research Center for International Economics, University of Washington

    brain drain

Jon Wongswan

Emerging Markets Researcher, International Active Equities Group, Barclays Global Investors, San Francisco

    spillovers

Kevin Honglin Zhang

Associate Professor of Economics, Illinois State University

    foreign direct investment and export performance

Laixun Zhao

Professor, Research Institute for Economics and Business, Kobe University

    unions and foreign direct investment

◼ absolute advantage

A country is said to have an absolute advantage over another country in the production of a good or service if it can produce that good or service (the "output") using fewer *real* resources (like capital or labor, the "inputs"). Equivalently, using the same inputs, the country can produce more output. The concept of absolute advantage can also be applied to other economic entities, such as regions, cities, or firms, but we will focus attention on countries, specifically in relation to their production decisions and international trade flows. The fallacy of equating absolute advantages with cost advantages is a never-ending source of confusion. Deviations between the two are caused by the fact that real resources may receive different remunerations in different countries.

In reaction to the mercantilist literature of the 17th century (which advocated state regulation of trade to promote wealth and growth), a doctrine of free trade emerged at the end of the 18th century, culminating in 1776 in Adam Smith's masterpiece, *An Inquiry into the Nature and Causes of the Wealth of Nations*. Drawing on the work of others, Smith was able to put many different arguments and elements together in a coherent and systematic framework, organized using a few general principles, and thus providing a new way of thinking about political economy (Irwin 1996). Smith thus provided the first *analysis* of economic reasons for advocating a policy of free trade and, according to Joseph A. Schumpeter (1954, 374), "seems to have believed that under free trade all goods would be produced where their absolute costs in terms of labor are lowest."

Smith's arguments can be summarized as follows. First, he points out that regulations favoring one industry draw away real resources from another industry, where they might have been more advantageously employed (opportunity costs). Second, he applies the opportunity cost principle to individuals in a society   for example, by pointing out that the tailor does not make his own shoes (which would cost him a lot of time) but buys them from the shoemaker (who can produce them more efficiently). Each individual is therefore specializing in the production of those goods and services in which he or she has some advantage. Third, Smith applies the same principles of opportunity costs and specialization to international commercial policy and nations. It is better to import goods from abroad where they can be produced more efficiently, because this allows the importing country to focus production on the goods it can itself produce efficiently. The primary (classical) reason for international trade flows is therefore a difference of technology between exporter and importer.

Principle of Absolute Advantage To illustrate the principle of absolute advantage, suppose that there are two countries (the United States and Japan) producing two goods (food and cars), using labor as the only input. Assume that goods can be traded without costs and workers are immobile between the two countries, but mobile between the two sectors within a country. All workers in a country are equally productive. Production technology in Japan differs from that in the United States (see table 1). We assume that Japan requires three units of labor to produce one unit of food, whereas the United States

Table 1
Productivity tables, an example of absolute advantages

	a. Units of labor required to produce one unit of output		b. Units of output produced with one unit of labor	
	food	cars	food	cars
USA	2	8	1/2	1/8
Japan	3	6	1/3	1/6

requires only two units of labor. Similarly, Japan needs six units of labor to produce one car, whereas the United States needs eight units of labor. Since Japan is more efficient in the production of cars and the United States is more efficient in the production of food, Japan has an absolute advantage in the production of cars and the United States has an absolute advantage in the production of food.

To show that specialization of production, coupled with international trade flows according to absolute advantage, can be advantageous, in our example suppose that the United States produces one car less. This frees up eight units of labor, which can now be used to produce $8/2 = 4$ units of food (opportunity cost of car production in the United States). The United States has now produced one car less and four units of food more. Suppose that the United States wants to consume the same number of cars as before. It must then import one car from Japan. To produce this car Japan needs six units of labor. These laborers must come from the food sector, where production therefore drops by $6/3 = 2$ units of food (opportunity costs of car production in Japan). Now note that the total production of cars has been unchanged (one car less in the United States and one car more in Japan), while the total production of food has increased by two units (four units more in the United States and two units less in Japan). These extra units of food reflect the potential gains from specialization if both countries concentrate in the production of the good they produce most efficiently. In principle, both countries can gain: for example, if they exchange three units of food for one car.

Complications and Limitations There are several caveats to the foregoing analysis, some of which we discuss now.

Absence of absolute advantage: The example discusses a situation where one country has an absolute advantage in the production of one good and the other country in the production of another good. It is frequently argued that developing countries may lack the technology to gain an absolute advantage in the production of *any* good, such that they cannot possibly compete on the global market and benefit from free trade (in table 1, for example, if the United States needs four laborers to produce one unit of food). This conclusion is wrong, however, according to David Ricardo's model of comparative advantage (which emphasizes labor as the primary production factor and attributes the costs and benefits of trade to the differences in opportunity costs among countries), since technologically disadvantaged countries can compete on the global market by paying lower wages. It turns out that absolute advantage is neither a necessary nor a sufficient condition for exporting a certain good and gaining from international trade.

More factors of production: In reality, goods are produced using several factors of production simultaneously, such as capital, land, and various types of labor. Usually, goods then cannot be ranked according to absolute advantage as their production in one country requires more of one input and simultaneously less of another input than in another country. These issues are analyzed in the Heckscher-Ohlin (factor abundance) theory of international trade.

Intra- versus interindustry trade: The example discusses interindustry trade, which is the exchange of one type of good (cars) for another type of good (food). Many countries engage in intraindustry trade, the exchange of similar types of goods (e.g., simultaneously exporting and importing car parts). This type of trade is becoming ever more important. It can be based on market power and economies of scale, as analyzed in New Trade Theory.

Absolute Advantage, Income, and Wages Despite the limitations and complications just discussed, absolute advantages (as reflected by differ-

ences in technology) *are* important for explaining current international trade flows and differences between countries in terms of income levels and wage rates. Daniel Trefler (1995) systematically analyzes these issues by combining the Heckscher-Ohlin model with technology differences, while taking into consideration the empirically observed home country bias (a consumer preference for domestically produced goods over otherwise identical imports). This combination explains about 93 percent of international trade flows. It also shows that technology differences are largely responsible for the deviations in income levels (and wage rates) between, say, the African countries and the high-income countries of the Organisation for Economic Co-operation and Development. For this reason absolute advantage does retain relevance for understanding the modern world economy.

See also comparative advantage; economies of scale; gains from trade; Heckscher-Ohlin model; intraindustry trade; new trade theory; revealed comparative advantage; Ricardian model; trade and wages

FURTHER READING

Irwin, Douglas A. 1996. *Against the Tide: An Intellectual History of Free Trade*. Princeton, NJ: Princeton University Press. A magnificent overview of the arguments for and against free trade throughout history.

Schumpeter, Joseph A. 1954. *History of Economic Analysis*. 12th printing, 1981. London: Allen and Unwin. Still *the* history of economic analysis.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Edited by R. H. Campbell and A. S. Skinner. The Glasgow Edition of the Works and Correspondence of Adam Smith 2. Reprint, 1981. Indianapolis, IN: Liberty Press. The starting point of economics as a science, using a coherent system of analysis favoring free trade.

Trefler, Daniel. 1995. "The Case of the Missing Trade and Other Mysteries." *American Economic Review* 85: 1029 46. Ingenious empirical tests of various trade theories with a prominent role for technology differences.

CHARLES VAN MARREWIJK

■ **access to medicines**

The term *access to medicines* encompasses the array of problems faced by the world's lowest-income inhabitants, who often cannot afford, or do not have access to, medications that could greatly reduce the disease burden under which they suffer. The problems include deficient medical infrastructure, imbalances between prices and ability to pay, and the lack of incentive to develop medicines that would treat diseases endemic to low-income nations.

During the 20th century, numerous technological breakthroughs in pharmaceutical therapy made it possible to cure or at least alleviate most of the diseases that have killed or debilitated millions of people each year. But the ability to purchase those medicines is concentrated in relatively affluent nations, where the vast majority of pharmaceutical sales occur. At the other extreme, roughly 60 percent of the world's population live in nations defined by the United Nations in 2000 as "low income," with per-capita gross national product averaging less than $530 (at prevailing exchange rates) a year in 1998. The World Health Organization (WHO) (2004, 61) estimates that in 1999 those nations made only 2.9 percent of the world's pharmaceutical purchases. The WHO has predicted that by expanding access to available health interventions, and especially essential medicines, 10.5 million lives could be saved annually by the year 2015. Lack of access to medicines and complementary health care in turn perpetuates a vicious spiral: poor health impairs productivity and economic development, while low productivity keeps the citizens of the least-developed nations too poor to afford appropriate health care.

Affordability The medicine access problem has several facets. The overriding problem is inability of individuals to afford medicines. Health insurance is an absent corrective; an estimated 90 percent of the people in developing nations lack such insurance. Inability to pay restricts not only the demand for medicines but also the supply of physicians able to diagnose diseases and recommend appropriate therapies. Nations classified as low income in 1998 by the United Nations had 70 physicians per 100,000 population; those classified as high income, 252. In

many instances, the only advice available comes from traditional practitioners whose herbal remedies may work for some indications, but with at best erratic success due to the lack of evidence from controlled experiments. For diseases such as AIDS, malaria, and tuberculosis, carefully administered therapy regimens must be maintained to inhibit the emergence of resistant strains. Counterfeit versions of first-world drugs continue to be a significant problem, in part because third-world health-care authorities lack systematic testing and approval institutions.

In their efforts to combat the burden of disease, health authorities at the WHO and in individual less-developed nations have since 1977 published "model lists" of so-called essential drugs. Drugs have been included on the list in part because of their proven efficacy and partly because of their relatively low cost. Low cost in turn has been achieved by emphasizing generic drugs, that is, those on which patent rights restricting supply to a single firm have expired. Historically, more than 90 percent of drugs on the WHO's model lists have been generics. However, this emphasis was threatened by the emerging epidemics of HIV/AIDS and related opportunistic diseases such as resistant tuberculosis and cryptococcal meningitis. Virtually all of the drugs effective against those diseases were patented and, at least initially, available only at costs for a year's treatment exceeding total average incomes of citizens in low-income nations.

The Trade-Related Aspects of Intellectual Property (TRIPS) agreement culminating the World Trade Organization's (WTO) Uruguay Round Treaty, signed at Marrakech in April 1994 and implemented in 1995, exacerbated this situation. Up to that time, many less-developed nations, emulating some more prosperous countries, were able either to produce or, more likely, import patented drugs because they granted no patent rights on pharmaceutical product inventions. TRIPS required that signatories to the Marrakech treaty begin awarding such patents within one year for the wealthiest nations, five years for middle-income countries, and ten years (later extended to the year 2015) for the least-developed nations. Transitional provisions also re-

quired grants of marketing exclusivity for post-1994 inventions on which initial patent applications were filed. Especially for AIDS and AIDS-related diseases, this posed special problems. Up to that time, the newest and most effective drugs might be available generically from India and other nations that had not awarded pharmaceutical product patents. But as the TRIPS provisions began to bind on India, Brazil, South Africa, and other nations, their ability to continue supplying low-cost generics atrophied.

The combination of TRIPS and the AIDS epidemic precipitated a crisis. Two main solutions emerged. First, at a joint WHO WTO conference in Høsbjør, Norway, in April 2001, a consensus emerged encouraging the world's leading research-oriented pharmaceutical companies to practice "differential" or "Ramsey" pricing. The companies would charge high prices in rich nations and make life-saving drugs available to consumers in low-income nations at prices approaching marginal cost. From what had been near parity of AIDS drug prices across rich and poor nations (Scherer and Watal 2002), wholesale prices were shown by Lucchini et al. (2003) and the UK Department for International Development (2005, 22) to have plummeted in the least-developed nations, in some cases by as much as 98 percent. One consequence of such discriminatory pricing was the reexport of low-price drugs to high-price nations, but steps to suppress this "parallel trade" were quickly implemented. Donations from multinational pharmaceutical firms to organizations providing health care in less-developed nations in effect, sales at a zero price also helped increase access to essential medicines.

Second, because of exceptions written into the original TRIPS agreement, nations were able to threaten or actually implement compulsory licensing of existing or new patents on AIDS and other epidemic disease drugs in order to authorize generic production. Threats of compulsory licensing induced multinational patent holders to reduce sharply the prices of their branded drugs in the third world and enter into voluntary agreements with such nations as Brazil and South Africa to permit generic supply. A limitation in the original TRIPS text was

that production under a compulsory license was to be "predominantly for the supply of the [Member's] domestic market." However, many of the world's least-developed nations lacked both the technological know-how and sufficient market scale to produce generics for their own use. A permissive amendment to the TRIPS agreement accepted in August 2003 following a mandate issued at the Doha Round of international trade negotiations in 2002 alleviated this problem. The TRIPS agreement does not require nations formally to report compulsory licensing decrees, and as of 2006, only an AIDS drug license by Thailand to a government entity, minimally controversial under TRIPS, had come to light publicly. The existence of other unreported cases cannot be ruled out. Alternatively, post-2000 price and voluntary license developments may have been sufficient to satisfy the limited ability of low-income nations to distribute drugs effectively.

Incentives for Drug Development Another fundamental problem preventing access to medicines is the lack of innovative drugs targeted specifically toward diseases prevalent only in the third world, for instance, sleeping sickness, Chagas disease, and leishmanisais. Because low-income nations have limited purchasing power, multinational pharmaceutical firms lack demand-based incentives for research and testing on drugs targeted toward the so-called tropical diseases and the resistant strains that continue to evolve. A study for Medicins sans Frontières (2001) revealed that among 1,393 new drug chemical entities introduced into world markets between 1975 and 1999, only 13 (or 15 counting tuberculosis) drugs were indicated for tropical diseases. Also deficient has been the development of vaccines that could *prevent* diseases curable using modern medicines, but at costs too high to be sustained by overstressed third-world medical care providers.

Here too the AIDS crisis played an important role in inducing corrective initiatives. Some large multinational pharmaceutical companies, seeing the problem as a moral challenge, increased research and development (R&D) efforts targeted at third-world diseases and established new laboratories nearer the potential markets. Private philanthropic organiza-

tions such as the Gates Foundation have provided generous subsidies to support R&D on new drugs and vaccines to combat third-world diseases. Their efforts complemented the work of the UN AIDS initiative and similar programs by national governments. In 2005 6, delegates from the world's eight largest market economies (the G-8) approved in principle a program to stimulate the development of vaccines by agreeing to purchase at generous pre-specified prices $3 billion worth (in each category) of new vaccines effective against AIDS, malaria, and tuberculosis. However, as of 2006, the G-8 member nations were tardy in backing their good intentions with actual purchase guarantees and the national budget commitments necessary to implement them.

Progress is being made in increasing the supply of affordably priced medicines to low-income nations, but much remains to be done. Overcoming the remaining barriers to access to medicines could alleviate disease worldwide and contribute to economic development.

See also Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS); health and globalization; HIV/AIDS

FURTHER READING

Granville, Brigitte, ed. 2002. *The Economics of Essential Medicines.* London: Royal Institute of International Affairs. A comprehensive collection of relevant articles, with emphasis on the consequences of TRIPS.

Levine, Ruth, Michael Kremer, and Alice Albright. 2005. *Making Markets for Vaccines: Ideas to Action.* Wash ington, DC: Center for Global Development. In vestigates thoroughly how nations could join together to offer advance market commitments providing incentives for the development of new vaccines.

Lucchini, Stephane, et al. 2003. "Decrease in Prices of Antiretroviral Drugs for Developing Countries: From Political 'Philanthropy' to Regulated Markets?" In *Economics of AIDS and Access to HIV/AIDS Care in Developing Countries*, edited by Jean Paul Moatti et al. Paris: Agence de Recherches sur le Sida, 170 211. Traces fall of AIDS drug prices from 1997 to 2002.

Medicins sans Frontières, Access to Essential Medicines Campaign and Drugs for Neglected Diseases Working

Group. 2001. *Fatal Imbalance: The Crisis in Research and Development for Neglected Diseases*. Geneva: Medicins sans Frontières (September). A careful study of deficient incentives for R&D on diseases prevalent mainly in poor nations.

Scherer, F. M., and Jayashree Watal. 2002. "Post TRIPS Options for Access to Patented Medicines in Developing Nations." *Journal of International Economic Law* 5(4): 913 39. Prepared for the Høsbjør conference, explores theoretically and empirically the pricing of AIDS drugs in third world nations and explores income tax incen tives for drug donations.

UK Department for International Development. 2005. *Increasing People's Access to Essential Medicines in Devel oping Countries: A Framework for Good Practice in the Pharmaceutical Industry*. London: UKDID (March). Reviews the broad access problem and summarizes UK initiatives to relieve the problem.

World Health Organization. 2004. *The World Medicines Situation*. Geneva: WHO. An excellent overview with statistical data.

F. M. SCHERER

■ **affiliate**

See foreign direct investment (FDI)

■ **African Caribbean Pacific–European Union (ACP-EU) partnership agreements**

The trade and development relationship between the European Union (EU) and the African, Caribbean, and Pacific (ACP) countries has been shaped by a number of formal treaties and agreements since the end of World War II. The aim of these agreements has been to promote, with EU participation, opportunities for growth and development among the ACP countries by both direct and indirect policy measures. The more direct methods have included development funds, investment loans, and compensatory payments, while indirect methods have centered on trading arrangements and protocols that favor exports from ACP countries in a bid to generate growth. Underpinning these specific aspects, though, has been a more general focus on engendering wider social, political, and economic development as part of an understanding (often referred to as the *acquis*) between the EU and the ACP countries, which reflects the fact that the relationship is not only a trading club.

The Treaty of Rome, which established the European Economic Community (EEC) in 1957, included a section entitled "The Association of the Overseas Countries and Territories." This made specific provisions for the relationship between the EEC and the overseas territories and former colonies of member states under Articles 131 to 136. The association with former colonies of the six members of the EEC was designed "to promote the economic and social development of the countries and territories and to establish close economic relations between them and the Community as a whole" (Article 131, Treaty of Rome). In practice, these arrangements had their greatest effect on the former French colonies in West Africa and the Caribbean. Preferential trading arrangements formed the major part of the association with a commitment to review the policy after five years.

The Yaoundé Conventions The first review produced a new set of arrangements embodied in the first Yaoundé Convention (or Yaoundé I) signed on July 20, 1963, in the Cameroon capital by 18 countries of the Association of African States and Madagascar (AASM) and the six EEC states. Yaoundé I aimed to encourage the development of the AASM countries mainly by allowing preferential treatment of their manufactured exports into the EEC, but with only limited preference for agricultural exports. In return, the EEC was permitted to export limited volumes of manufactures to the AASM with similar duty arrangements. In addition to trade provisions, there was also agreement on technical and financial issues, on rights of establishment that allowed for, among other commercial features, the establishment of companies in associated states, and also on the institutions that would oversee the governing of the convention. The agreement ran from 1964 to the end of 1969.

Countries that were not part of the Yaoundé I sought associate status. Under the Arusha Agreement signed in Tanzania on September 24, 1969, Kenya, Tanzania, and Uganda negotiated associate status with the EEC. This came into force at the same time as the second Yaoundé Convention (Yaoundé II), which was in effect from 1971. While reenforcing the preferential and reciprocal trade arrangements, Yaoundé II also included provision for investment by the EEC in the associated states. Specifically, funds were provided mostly for the European Development Fund (EDF) with a small amount going to the European Investment Bank (EIB) for loan-supported project work. The specific aim was to broaden the relationship between the two groups, from trade policy to wider development areas. The Arusha Agreement only contained trading elements and none of the financial aid offered under Yaoundé II.

From Yaoundé to the ACP In 1975, the developing country signatories to the Yaoundé Conventions formed a new alliance with the 20 Commonwealth countries associated with the United Kingdom (UK). The new body was called the African Caribbean and Pacific (ACP) Group. The terms of this new body were established within the Georgetown Agreement. The main aim was to coordinate negotiations for ACP countries with the EEC, a process that had begun in 1973 as part of a review of Yaoundé II. The negotiations were concluded with the signing on February 28, 1975, of the first Lomé Convention (Lomé I) in Togo, by 46 ACP countries and the then nine EEC Member States.

Lomé I had a number of provisions but the key ones again related to trade. Free access for most ACP exports to the EEC, although unlike Yaoundé II without reciprocal terms, lay at the heart of it. In addition, the agreement introduced specific protocols for sugar, rum, bananas, and beef and veal. In the Sugar Protocol, for example, the EEC agreed to volume import quotas of raw (cane) sugar from ACP producers at a guaranteed minimum price. The protocol reflected the UK's entry into the EEC and its established trading agreements with its former colonies. ACP sugar producers were allocated quotas for exports with the aim of aiding their producers

without harming EEC producers of beet sugar. The other commodities had similar export quota and guaranteed price arrangements, although beef and veal saw refunds of tax at 90 percent on imports.

In addition to these trade arrangements, the convention also provided for a Council of Ministers. This body was drawn from members of the Council and Commission of Ministers for the EEC and representatives from each ACP country, with the presidency alternating between the two groups. The other significant innovation was a change in the nature of EDF financing. The STABEX (shorthand for stabilization of export earnings) scheme aimed to provide stabilizing finance when export earnings fell due to a decline in prices for a producer's main (often primary) exports. This reflected the concerns about volatility in world commodity prices and the impact on exporters and countries' macroeconomic planning and policies. Coupled with further EDF and EIB monies, the convention moved explicit financial aid more prominently into the relationship between the EEC and the ACP, albeit with a continued emphasis on expenditure on infrastructure.

Lomé II was agreed and signed in 1979. Although it did not offer new trading provisions, within its EDF provisions it did introduce SYSMIN (stabilization of export earnings from mining products), a system of loans for helping the mining industries in those countries that relied heavily on exports of minerals for revenue generation, to diversify into other sectors. Lomé III (1984) signaled a shift from direct encouragement of export-led growth to encouragement of self-sufficiency and especially security of food supplies. Rural development was promoted as a means of achieving these goals. Finally, Lomé IV (1990) covered a 10-year period with a five-year review of financial support. However, it also became apparent that wider social issues, such as the environment, women's roles, and diversification of the economy were given much greater prominence as the ACP countries continued to develop. The EU recognized a desire for greater self-determination of policy.

A major review of the Lomé Convention came in 2000. The Cotonou Agreement of 2000 was signed

in Benin and built on the Lomé acquis but took a new, longer-term approach to the political, trade, and development aspects of ACP-EU relations. Globalization had appeared to pass many ACP countries by, with their share of foreign investment flows being very small and their trade shares equally limited. Tied to decreasing donor aid, this presented a problem that the Lomé IV had not dealt with. Indeed, compliance with World Trade Organization (WTO) rules meant protective trade arrangements could no longer provide an answer even if they were desired. Instead, focus on poverty reduction via good governance, macroeconomic stability, and new trading arrangements increased. To integrate ACP countries more fully with global markets, the EU liberalized virtually all imports from least-developed countries (LDCs), not just ACP countries, under a General System of Preferences (GSP). The protocols for sugar and beef and veal remained, however. Funding was now via grants totaling 11.3 billion euros and for risk capital, which totaled 2.2 billion euros.

In 2001 the EU concluded its amendment of the GSP and developed its "Everything but Arms" policy. This policy extended duty-free access to all LDC exports apart from arms and munitions, with some restrictions still applying over a longer period for bananas, rice, and sugar. Of the 48 LDCs, 39 were ACP (Cotonou signatory) countries.

The WTO continued to put pressure on the EU to move away from preferential treatment of ACP exports, and in 2002 Economic Partnership Agreements (EPAs) became the focus for ACP-EU trading relationships. The negotiations with regional groupings sought to encourage partnership, regional integration, development, and ultimately integration of the ACP countries into the WTO. EPAs were scheduled to be in place by 2008.

Given the scope, scale, and relative complexity of the various ACP-EU agreements, it is possible to view them as central to a continuing process of different countries working together for mutual benefit. Although not comprehensive in either geographic or economic coverage, ACP-EU agreements have played a major role in shaping trading policies for many countries and have offered possible options for others to follow.

See also European Union; international trade and economic development; World Trade Organization

FURTHER READING

European Commission External Relations Directorate. http://europa.eu/scadplus/leg/en/s05032.htm. This site provides a wealth of materials relating to the establishment and development of EU ACP trade agreements as well as links to other helpful documents.

Secretariat of the African, Caribbean and Pacific Group of States. http://www.acpsec.org/. The ACP site offers a number of useful resources that provide greater detail on the trade agreements with the EU, as well as giving an overview of the structure of the ACP group and how it operates.

C. W. MORGAN

■ African Development Bank
See regional development banks

■ African Union

Open to all countries in the African continent, the African Union (AU) is an organization designed to foster political and economic cooperation and development among its member countries. To such ends, it stands ready to address any and all issues relevant to state building, security, and economic development and integration among countries on the African continent. Hence the AU can contribute to factors deemed essential to greater integration of the continent in the world economy. It was officially launched on July 9, 2002, replacing the Organization of African Unity (OAU), whose charter was signed on May 25, 1963, with an original membership of 33 countries; the AU has 53 members. The headquarters are in Addis Ababa, Ethiopia, although the various organs can be located in other member states; for example, the Pan-African Parliament is in Midrand, South Africa.

Political Stability and Security In order to advance political stability and security, the AU focuses on conflict resolution within and between states, peer review among the African states to facilitate state building and the democratization process, and building solidarity to increase the leverage exercised by African countries at the international level. Still, the AU has found it difficult to speed up democratic transition in the continent; impediments to this transition include the manipulation of institutions by elites or breakdowns in the democratic political process because of ethnic conflicts or political fragmentation. The AU has also been handicapped in dealing with the resolution of conflicts in which the sources of conflict are deep seated and the combatants well armed.

With few exceptions, the AU has supported the territorial integrity of the African states since independence from colonialism, as well as noninterference in the internal affairs of those states. Enshrined in the Constitutive Act of the AU are the "condemnation and rejection" of "political assassinations," "subversive activities," and "unconstitutional changes of governments." Moreover, the AU has pronounced resolutely in favor of human rights. Thus one of the tenets of the Constitutive Act is the "right of the union to intervene in a Member State pursuant to a decision of the Assembly in respect of grave circumstances, namely war crimes, genocide and crimes against humanity."

Economic Integration The economic integration program of the AU is contained in the June 1991 Treaty Establishing the African Economic Community (AEC) signed in Abuja, Nigeria. That treaty has been operational since May 1994. The plan contained in the treaty calls for the AEC to reach fruition after a period of 34 to 40 years from 1994. The consequence, among other things, would be a single domestic market and a Pan-African Economic and Monetary Union, a single African Central Bank, and a single African Currency. A number of regional economic communities (RECs) operate under the aegis of the AEC, as part of the transition to full, continentwide union, namely, the Arab Maghreb Union (AMU), the Economic Community of the Central African States (ECCAS), the Common Market for Eastern and Southern Africa (COMESA), the Southern African Development Community (SADC), and the Economic Community of West African States (ECOWAS).

Progress in economic integration has been hampered by certain political and economic strains, overlapping membership among the RECs, competing subregional groupings within RECs, and a lack of clear commitment to integration among the populations and the political leadership. Political difficulties have included personal animosity among heads of states and governments; ideological differences among leaders; deep-seated disputes such as that over the Western Sahara (independence for a Sahrawi Arab Democratic Republic), in the case of the AMU; and regional conflicts, as in the Great Lakes area for ECCAS. But prospects are improving in these respects: Increasing democratization and acceptance of market solutions to economic problems are reducing ideological differences. Conflicts involving many states simultaneously are diminishing in number and those that remain are being better handled by the AU. Moreover, proponents of integration have been working hard to ensure that institutions and organizations of regional economic communities can function in spite of temporary personal hostilities in high political circles.

There is need to rationalize membership of RECs by encouraging countries to join only one. Also, subregional organizations with the same goal of economic integration exist. The best example is the East African Community (EAC), the three members of which are also members of either COMESA or SADC. In addition, the Francophone African countries are apparently happy with their monetary union arrangements. But they have been expanding their cooperation objectives in the direction of general economic integration, despite their membership in ECOWAS and ECCAS.

Economic obstacles to integration include (1) fear of a loss of national sovereignty over macroeconomic policy to some union authority or body; (2) disagreements over the nature and content of protection of local industries through tariffs and nontariff

barriers, which reduce certain imports of commodities and services from outside an REC; and (3) concern about unequal distribution of gains and losses of REC membership. The RECs continue to make progress toward resolving these issues.

For instance, a common external tariff is an important objective of the RECs, and the structure of a tariff system has important implications for the protection bestowed on different commodities. If a simple rule were established, such as equal protection for all commodities, then the determination of tariff rates could be left to technical experts to decide. But to assist infant industry and foster industrial development, African countries want differential protection. Given the economic structure and the state of development of the various countries concerned, different schedules of tariff rates have dissimilar implications for comparative advantage of the countries. Hence, such considerations seriously affect discussions of the detailed tariff schedules to be put into effect.

Many in the continent fear that gains to countries from economic unions will be positively related to the degree of their economic development and/or the size of their domestic economies. The allegedly "unfair" distribution of gains and losses is widely believed to have been at the root of the breakup of the first EAC, where it was felt that Kenya's industrialization was greatly helped but, in the process, Tanzania's may have been adversely affected. An attempt to use differential intraunion tariffs designated transfer taxes could not alleviate the problems, at least not to the satisfaction of Tanzania.

In general, many want some kind of internal (intraunion) tariff structure that protects some national domestic production activities from direct competition within an REC. But once the principle is accepted (and applied) that the location of industries among countries should be determined in a world of open competition and free mobility of all factors of production including labor, rather than in an arena of negotiated industrial planning buttressed by restricted mobility of factors, especially labor, the case for transfer taxes becomes weak.

A challenge would still remain as to how to balance such a market-oriented approach to the location of industries with permitting selective intervention of governments for economic development of the countries, as deemed useful by all the countries. The difficulty would be compounded by the need to observe certain macroeconomic constraints set by the union as a whole for example, limits on government budget deficits and on government debt in relation to gross domestic product.

Differences in taxation systems and structures also continue to engender issues of unequal gains and losses. In particular, countries have different reliance on import taxes as sources of government revenue. This fact has slowed down reduction of intraunion tariffs, since a formula to compensate those who will lose tax revenue from large intraunion tariff reductions is not easy to negotiate. Thus countries realize they need to reform their tax systems to lessen their dependence on import taxes if substantial and rapid intraunion tariff reductions are to occur in practice. The attempts of countries to reform their tax systems and to move toward greater reliance on income, profits, and value-added taxes should be of help in this regard.

One theme in the integration debate in the African continent is the degree to which African leaders are committed to full economic integration in the foreseeable future. For many of the countries, intraregional trade is very small in relation to extraregional trade, and the countries in each of the regions often produce similar goods. Hence countries sometimes do not feel an urgent need for a common market, given the widespread belief that integration would not yield substantial economic benefits for some time.

Still, every single leader of the countries voices the view that, in time, the benefits of integration will be substantial, as the effective size of domestic markets will greatly enlarge, so that technological economies of scale can be realized and the returns to investment enhanced. Hence it is along these lines that the most fervent proponents of integration have argued their case. Those who prefer a slower pace are content to

push now for (1) promotion of greater intraregional trade, employing the instrument of a common external tariff, probably supplemented by some form of transfer tax or *taxe de coopération regionale*, until full labor mobility becomes socially and politically feasible; (2) cooperation in infrastructure and industrial "regional" projects; and (3) some harmonization of policies (especially macroeconomic) as feasible. This could be followed, in the eyes of the gradualists, by some form of monetary union. Only later, when mass support for integration is strong and ideological obstacles are minor, according to this perspective, should full integration be pursued.

Governance and the African Peer Review Mechanism The AU aims at improving *governance* in African countries, in a context of enhanced country *ownership* of policymaking. In 2001, the AU launched the Millennium Partnership for the African Recovery Program (MAP). It was billed as a pledge by African leaders to take decisive steps to improve governance, reduce poverty, and enhance economic growth of their countries. In particular, it claimed that a new crop of leaders was emerging in Africa committed to democracy and the integration of their countries into the world economy. It called for "a new relationship" with the international community, especially the industrial countries: African countries would take charge of their own destiny, and the rest of the international community was called on to make a concerted effort to enhance resource flows to the continent via "improvements" in aid, trade, and debt relationships. Several goals were specified, including most notably achieving a 7 percent average annual growth rate of gross domestic product over the following 15 years. Among the "policy thrusts" to achieve the objectives would be negotiating "a new partnership" with the industrialized countries and multilateral organizations. African "ownership, leadership, and accountability" were thus highlighted as central elements of the MAP. The African peoples were henceforth going to set and direct their agendas and shape their own destinies. This, then, is the idea of the New Partnership for Africa's Development (NEPAD).

Within the NEPAD framework, the African countries have instituted the African Peer Review Mechanism (APRM). Participating countries will do self-assessments, using the services of domestic autonomous bodies and individuals who in turn involve business and civil society groups throughout the countries. The governments will then draw up programs of action to address weaknesses identified in the self-assessments in the areas of political governance, economic governance, corporate governance, and socioeconomic governance. Review teams of African experts will visit the countries to assess the integrity of the self-assessment exercise and make recommendations, including on the action plans of the governments. Future expert teams will visit to review progress in implementing the action plans. Central in this arrangement will be a panel of eminent persons of the continent, overseeing the APRM processes to ensure their integrity and guiding the preparation of the country reports drafted mainly by the experts to be presented to the African Peer Review Forum. This forum comprises heads of state and government of participating countries (the "peers").

If high and transparent standards are maintained, the APRM can be an effective means of separating those African countries committed to good policies from the rest, because only those countries whose leaders are committed to implementing good policies will want to have their progress continuously reviewed and made known to the global community. In this respect, the APRM could address a major credibility problem: Africa as a region is considered high-risk for investors, and the credit ratings of countries within the region are adversely affected simply by their being there. The APRM can contribute to separation of African countries into those with good policy environments and those without. In addition, if the reports get widely circulated within the continent, and especially in those countries that have chosen not to participate, the APRM will help provide essential information to potential actors in civil society.

Moreover, if the APRM is to have any effect on NEPAD, and especially influence the aid and debt

relationships, it would be important that it become credible among aid donors, who then allow it transparently to influence their aid policies. Those of the international community interested in providing aid to support good policies may want to see evidence that the APRM is influencing governance in the right direction, for recent research in the social sciences has concluded that good policies emerge exogenously, when countries own such policies and voluntarily adopt them.

Future of the African Union The AU is poised to have an enhanced, though still limited, role in the world economy in the foreseeable future. Its efforts are bringing peace, political stability, and democratization to African states. Peace and political stability are good for economic growth and democratization improves governance. But good political leadership remains elusive and this ultimately is the route by which the political regime has its greatest influence on economic growth. The AU is not likely to have much influence on political leadership in individual African countries. Economic integration also will proceed more slowly than envisaged by official AU pronouncements. But economic cooperation will accelerate, leading to faster infrastructure development, policy harmonization within the regional economic communities, and more efficient and development-oriented industrial, agricultural, and service projects. Moreover, in arenas such as the International Monetary Fund, the World Bank, and the World Trade Organization, African countries more frequently will speak with one coherent voice under the aegis of the AU.

See also Common Market for Eastern and Southern Africa (COMESA); Economic Community of West African States (ECOWAS); European Union; regionalism

FURTHER READING

African Union. 2003. "African Peer Review Mechanism: Organization and Processes." Addis Ababa, Ethiopia: African Union. Downloadable from http://www.nepad.org/2005/files/documents/48.pdf. Describes the APRM processes.

El Agraa, Ali M. 2004. "The Enigma of African Economic Integration." *Journal of Economic Integration* 19 (1): 19

45. The paper critically examines the aims of the AU, arguing that they were not carefully thought out.

Johnson, Omotunde E. G. 1995. "Regional Integration in Sub Saharan Africa." *Journal of European Integration* 18 (2 3): 201 34. Discusses the central themes of debate on economic integration in sub Saharan Africa.

———. 2004. "How Will Good Economic Policy Environments Emerge in Africa?" *Journal of Policy Reform* 7 (3): 151 64. Argues that the underlying problem of the policy environments in Africa can be found in weak societal demand for good economic policies and a lack of exogenous constraints to make it in the self interest of political elites to pursue good policies even without domestic demand pressures.

Murithi, Timothi. 2005. *The African Union: Pan African ism, Peacebuilding, and Development.* Aldershot, UK: Ashgate. Analyzes the challenges faced by the African Union in achieving its objectives, arguing that the African Union is another manifestation of Pan Africanism.

OMOTUNDE E. G. JOHNSON

■ **agglomeration**

See New Economic Geography

■ **agglomeration and foreign direct investment**

The spatial clustering of foreign direct investment (FDI) is clearly visible in the location of multinationals investing in the United States, the European Union, China, and other regions. This agglomeration is at least partly the result of policy, as in China's special economic zones, but spatial concentration is also characteristic of domestic firms and of FDI in economies with few controls. These observations suggest that market forces, as well as policy, lead to clustering.

That new establishments tend to go to the same locations as earlier entrants suggests that productivity rises with the level of economic activity, especially as firms often must pay higher land prices to locate in clusters. If such productivity-enhancing effects, or

agglomerative economies, exist and spill over to domestic activities, a case may be made for government incentives to multinationals to induce local affiliate production. Indeed, dozens of countries favor FDI through tax breaks and subsidies. Through these incentives, governments hope to begin a self-reinforcing process whereby subsidized early entrants attract additional investment.

To better design such policies, researchers have sought evidence that agglomerative economies exist and, if they do, the extent of their benefits to local productive factors. Location-choice studies seek to measure the attractiveness of local characteristics for foreign investors and thus provide a way to estimate the self-reinforcing power of FDI. Virtually all location-choice studies find that the existing stock of foreign investment is a significant predictor of the location a multinational will choose for new local affiliates. However, most countries receive a relatively small number of new multinational affiliates in a given year and for these projects there is often limited information, constraining our ability to identify the specific sources of agglomerative economies.

Head and Ries (1996) observe a relatively large number of investment projects, 931 equity joint ventures in 54 Chinese cities from 1984 to 1991. Their study is noteworthy for its careful modeling of the agglomerative process, emphasizing local input sharing as the source of positive firm spillovers. Using conditional logit analysis to estimate the likelihood that a particular city is chosen as the investment site, Head and Ries find that agglomerative economies greatly magnify the direct impact of government incentives. Their simulation analysis suggests that two-thirds of the gains from incentives can be attributed to the self-reinforcing nature of earlier investments. Not all locations gained equally, however, as cities considered attractive for other reasons, such as infrastructure and industrial base, gained the most. Similarly, Devereux, Griffith, and Simpson (2007) find that firms are less responsive to government subsidies in areas where there are fewer established plants in their industry.

Evidence that past investment increases the likelihood of new investment does not necessarily imply the existence of agglomerative economies. Agglomeration arises because there are benefits to locating near similar firms and because certain locations have natural advantages features of a location that are independent of firm location decisions. A common example of how natural advantages influence location choice is the North American steel industry, which concentrated in the Great Lakes region largely because of the location of iron ore and coal deposits. In measuring the extent of agglomerative economies, researchers confront an identification problem: Are firms choosing a common location because its inherent characteristics make them more productive or are they more productive because they have all chosen the same location?

Head and Ries (1996) try to separate the roles played by natural advantages and agglomerative economies in two ways. First, they include in their logit analysis a set of variables that attempt to control for local characteristics that influence firm productivity, particularly infrastructure. Second, they allow for spatially correlated errors by including provincial fixed effects. These two approaches are standard in the literature, and data limitations often make it difficult to do more to avoid bias caused by omitted local characteristics or endogeneity. For example, it is often impossible to include fixed effects at the same geographic scale as the unit of location choice (e.g., city fixed effects in the Head and Ries study) because they cannot be estimated for regions that received no investment. However, to fully control for all features of a location that attract investment is impossible, and even in the most careful studies omitted variables likely remain a problem.

Some studies have tried to assess the relative attractiveness of various kinds of prior investment for new entrants. Examining Japanese investment in the United States electronics industry from 1980 to 1998, Chung and Song (2004) ask whether firms agglomerate with their competitors or with their own prior investments. They find that firms tend to colocate only with their own prior investments, with the exception of firms that have little of their own experience, who do tend to colocate with competitors.

More recent work emphasizes the role of trade costs and market access as an alternative explanation for FDI clustering. Head and Mayer (2004) develop a theoretical model in which firms prefer to locate where demand is highest and serve smaller markets by exporting. They confront the data with this hypothesis, measuring market potential by a term that weights demand in all locations by its distance from the proposed investment site. Head and Mayer use standard logit techniques to analyze the European regions chosen as the sites of 452 Japanese investments. They decompose existing investment in each region into three firm counts distinguished by their relatedness to the new entrant: domestic establishments in the same industry, Japanese affiliates in the same industry, and Japanese affiliates with the same parent or network. They find that all three measures of prior investment have a large and positive influence on the likelihood that a region will be chosen by a new entrant, with this effect larger the closer the relations between firms. Thus there are strong agglomeration effects even when controls for market potential are included in the analysis.

An important issue for policy is whether domestic productivity is enhanced by the presence of foreign-owned firms. Most productivity-spillover studies are of specific industries or are case studies, both of which are limited as a guide to policy. Haskel, Pereira, and Slaughter (2007) offer evidence on domestic spillovers from FDI using a plant-level panel of all UK manufacturing firms from 1973 to 1999. Several previous studies using plant-level data find a negative or insignificant effect of industry-level FDI on local productivity. The UK data are unique in that they cover the whole of manufacturing in a developed country. Haskel, Pereira, and Slaughter estimate plant-level productivity and regress it on industry-level FDI, controlling for inputs and the level of competition. They estimate that a 10-percentage-point increase in foreign presence in a UK industry raises the total factor productivity of that industry's domestic plants by about 0.05 percent. They compare the value of these estimated spillover effects to per-job incentives offered in specific cases and find that these expenditures outweigh the benefits.

Haskel, Periera, and Slaughter (2007) use a variety of methods to deal with identification problems. In addition to explaining variation in gross output, they time-difference the data, explaining the change in output as a function of changes in inputs and foreign industry presence. This method accounts for plant-specific effects. The authors also include time, industry, and region fixed effects in their regression analysis. They also worry about the possibility that changes in industry FDI levels are correlated with changes in domestic productivity, and use instrumental variable techniques to minimize endogeneity bias. Their findings provide the strongest evidence to date that foreign investment does raise domestic productivity, but more work is needed before we have a clear guide to policy.

In sum, locations are more attractive the larger the existing stock of foreign investment, especially when the existing investments are by firms that are closely related (same industry, nationality, or parent firm). Government incentives are a significant determinant of multinational affiliate location choice but incentives are most effective when a location is desirable for other reasons. Although recent evidence suggests that foreign-owned firms enhance the productivity of local establishments, the value of these domestic spillovers appear to be less than the incentives used to attract foreign investment.

See also location theory; New Economic Geography; technology spillovers

FURTHER READING

Chung, Wilbur, and Jaeyong Song. 2004. "Sequential In vestment, Firm Motives, and Agglomeration of Japanese Electronics Firms in the United States." *Journal of Eco nomics and Management Strategy* 13 (3): 539 60. For the location choice of experienced firms, own prior invest ments matter more than competitors' investments.

Devereux, Michael P., Rachel Griffith, and Helen Simpson. 2007. "Firm Location Decisions, Regional Grants, and Agglomeration." *Journal of Public Economics* 91 (3 4): 413 35. Firms are less responsive to subsidies where there is little own industry investment.

Haskel, Jonathan E., Sonia C. Pereira, and Matthew J. Slaughter. 2007. "Does Inward Foreign Direct Invest ment Boost the Productivity of Domestic Firms?" *Re view of Economics and Statistics* 89 (3): 482 96. FDI raises domestic productivity.

Head, Keith, and Thierry Mayer. 2004. "Market Potential and the Location of Japanese Investment in the Euro pean Union." *Review of Economics and Statistics* 86 (4): 959 72. Prior investments matter more the closer the relations between firms.

Head, Keith, and John Ries. 1996. "Inter City Competition for Foreign Investment: Static and Dynamic Effects of China's Incentive Areas." *Journal of Urban Economics* 40 (1): 38 60. Agglomeration economies magnify impact of incentives.

Rosenthal, Stuart S., and William C. Strange. 2006. "Evi dence on the Nature and Sources of Agglomeration Economies." In *Handbook of Regional and Urban Eco nomics*, vol. 4, edited by J. V. Henderson and J. F. Thisse. Amsterdam: Elsevier, 2119 72. A detailed and clear exposition of the sources and nature of agglomerative economies and a thorough review of evidence based on domestic investment.

MARY E. LOVELY

■ Agreement on Agriculture

The Uruguay Round Agreement on Agriculture (URAA) came into effect in 1995 as a part of the Marrakesh Agreement that established the World Trade Organization (WTO). Contained in Annex IA of the Marrakesh Agreement, the URAA both modifies and greatly elaborates on those Articles of the General Agreement on Tariffs and Trade (GATT) that specifically dealt with agricultural trade by specifying significant constraints on government behavior in this area. The scope of the URAA covers all agricultural products (defined as products in Chapters 1 24 of the Harmonized System of tariff headings, excluding fish and fish products but in cluding cotton, wool, hides, flax, hemp, and a few other products as specified in Annex 1). The agree ment, by internal reference, also includes the country schedules that were appended to the WTO Treaty (Articles 3.1, 4.1, and 6.1). These schedules con tained maximum permitted levels for export sub sidies and for certain types of domestic subsidies, as well as commitments for the reduction of "bound" tariffs (tariff levels that cannot be exceeded without negotiating compensation for effected exporters).

The central elements of the URAA are often re ferred to as the three "pillars" market access, do mestic support, and export competition. In all three areas, new rules and reductions in trade barriers form a comprehensive framework for the regulation of measures that restrict trade in agricultural products.

Market access rules include the conversion of all nontariff import barriers (quotas and restrictive li censes) to tariffs (Article 4.2), and a footnote to Ar ticle 4.2 specifies some of the nontariff measures that are prohibited. Moreover, it was agreed that tariff levels were to be bound and that tariff rate quotas (TRQs, or quantities that can be imported at a zero or low tariff) were to be established to maintain market access as tariffication (replacement of nontariff bar riers with tariffs) took place. These TRQs were to represent "current access" in cases of existing trade or a "minimum access" of 3 percent of domestic con sumption (rising to 5 percent over the implementa tion period) in cases where there were no imports in the base period. Tariffs were to be reduced from the base period (1986 90) by an (unweighted) average of 36 percent, with a minimum cut of 15 percent for each tariff line, over a six-year period (1995 2000). In addition, the agreement established a special safeguard regime that countries could use to counter import surges or price drops in markets in which they had newly established tariffs (Article 5).

Domestic support was defined to include pay ments to farmers in addition to the transfers from consumers through border policies. These included deficiency payments, direct income supplements, administrative price systems, and subsidies for agri cultural research and government advisory programs for farmers for conservation compliance, and for other programs that benefited farmers directly. These elements of domestic support were put into three categories, which have become known as the Amber Box, the Blue Box, and the Green Box.

Amber Box measures were those tied to output or input prices or to current output levels. These were to be reduced by 20 percent (in aggregate) relative to the base period (1986 90) subject to de minimis amounts that were excluded from the commitment. The Blue Box contained subsidies that were tied to supply control programs: such subsidies were regarded as less obviously output-increasing. There was no reduction obligation for Blue Box policies, but such subsidies were restricted to payments based on fixed acreage and yield or paid on a maximum of 85 percent of production (Article 6.5). Green Box subsidies were defined (in Annex 2) as those unrelated to price and output ("decoupled"), which included research and extension, payments designed to compensate farmers for the cost of compliance with environmental regulations, and domestic food assistance programs. Both the general criteria (that they be provided from public funds and not act as price supports) and the specific criteria for each type of subsidy identified have to be met. Those subsidies that qualified as Green Box payments were not constrained, though they had to be notified by governments to the WTO Committee on Agriculture.

The domestic support commitments were implemented by means of a calculation of the Total Aggregate Measure of Support (Base AMS) (Article 6) for the base period. This included market price support given by administered prices (calculated by a price gap relative to a reference price), nonexempt direct payments, and other subsidies. Exemptions included the Blue Box and Green Box subsidies and a de minimis amount of 5 percent of the value of production for non-product-specific subsidies and 5 percent of the value of the output of an individual commodity for product-specific payments. The reduction commitments were applied to the Base AMS to give the annual commitment levels included in the country schedules, and each year the Current Total AMS is compared to this commitment.

The rules regarding export competition included a prohibition on new export subsidies (Article 8) and a reduction of existing subsidies by both volume and expenditure. A list of export subsidy practices that are covered is given in Article 9.1. Following the agreed modalities, country schedules were drawn up that provided for subsidy reductions relative to the base period of 36 percent by expenditure and 21 percent by quantity subsidized. In addition, rules were made more explicit with regard to food aid (Article 10.4), and countries agreed to negotiate limits on export credit guarantees (government underwriting of sales to purchasers that might lack creditworthiness) (Article 10.2).

To provide for "special and differential treatment" for developing countries, the level of reductions for tariffs and subsidies was set at two-thirds of that of developed countries, and the period of transition was extended from 6 to 10 years (i.e., 1995 2004). Developing countries were also allowed to exempt de minimis subsidies of up to 10 percent of product value for product-specific payments and 10 percent of total agricultural production for non-product-specific payments. In addition, certain additional categories of both domestic support (Article 6.2) and export subsidies (Article 9.4) were allowed. In the case of least-developed countries, no reduction commitments were required (Article 15.2). These least-developed countries are defined as the 48 countries eligible for World Bank/International Development Association assistance, and developing country status is self-declared.

In addition to the three pillars, the URAA mandated the formation of an Agricultural Committee (Article 17), charged with the monitoring of adherence to the agreement. Countries were to notify the committee in a timely fashion of their subsidy levels and any new subsidies that were introduced. Notifications have lapsed, however, and some major countries have not notified beyond the year 2001. The Agriculture Committee became the locus for new negotiations on the continuation of trade reform, meeting in special session.

In addition, the URAA provided a degree of shelter for domestic programs through a "Peace Clause" (Article 13) that limited the scope for the challenge of agricultural subsidies under the Agreement on Subsidies and Countervailing Measures. The Peace Clause was to operate for a period of three

years after the implementation period; it expired in 2003.

A further innovation in the URAA was the inclusion of a clause (Article 20) that mandated a continuation of the process of reductions in support and protection. To this end, there were to be new negotiations by the end of the period of transition (in effect, before 2000). Negotiations did indeed start in March 2000, and were incorporated in the Doha Development Agenda (DDA) at the Doha Ministerial in November 2001. The DDA talks were suspended in July 2006 and revived in January 2007.

The need for the development of new rules for agricultural trade in the Uruguay Round reflected both the unsatisfactory nature of the constraints incorporated in the GATT articles and the "disarray" that had characterized these markets for decades. The three GATT articles that had caused the most conflict were Article XI, which prohibits nontariff measures; Article XVI (as modified in 1955), which limits export subsidies; and Article XX, which permits the use of trade barriers in support of a range of domestic health and safety measures.

The part of Article XI that was considered unsatisfactory was the clause (Article XI.2(c)(i)) that allowed an exception to the prohibition of nontariff trade barriers in cases where the domestic production of an agricultural product was subject to supply control. Many countries had relied on this clause to restrict imports by quantitative trade barriers when domestic markets were being managed. As it was difficult to monitor the extent to which the domestic supply control was effective, exporters of the products concerned claimed that the import restrictions were in effect the dominant policy rather than just an adjunct to help reinforce the domestic production limits. Examples were quotas on Canadian dairy and poultry imports and those imposed by the United States under Section 22 of the Agricultural Adjustment Act (as amended), which mandated quantitative restrictions on imports of a number of goods when domestic programs were "materially interfered with" by imports.

Another complication related to Article XI was whether a "variable levy" (a tariff that changed frequently depending on the level of import prices, so as to stabilize domestic markets) was an "ordinary customs duty." If not, then it would have been constrained by Article XI. The European Economic Community (EEC, later the European Union, EU) had built its Common Agricultural Policy on such an import policy instrument. So the question as to whether the EEC was acting within the limits of the GATT was continually raised by exporting countries   though it was never resolved.

In the case of export subsidies, the problems revolved around the ambiguous nature of Article XVI. Though the original GATT article subjected both primary and manufactured product export subsidies to the same notification and consultation procedures, in 1955 it was agreed to add an explicit prohibition on export subsidies on manufactured goods. Agricultural export subsidies were constrained only by the obligation not to use such subsidies to capture "more than an equitable share" of world markets. Successive GATT panels failed to come up with a satisfactory definition of this concept, and agricultural export subsidies in effect escaped any discipline.

The problems that had arisen in the application of Article XX centered on the difficulty posed by the need to distinguish between those measures that were legitimate and effective regulations to protect against disease and those that were largely inspired by the desire to protect the economic interest of domestic producers. The clarification of Article XX was addressed by the Sanitary and Phytosanitary (SPS) Agreement, which was complementary to the URAA. By requiring risk assessment in the case of all health and safety regulations related to trade in plants and animals, the SPS Agreement created a greater degree of accountability. Regulations that are clearly motivated by economic rather than health protection can now be (and have been) challenged in the WTO.

The URAA has rendered the provisions in Article XI regarding supply control moot, as quantitative import restrictions are now prohibited. Similarly, the variable levy is explicitly included in the list of import barriers that are not allowed. By banning new export subsidies and including existing subsidies in

schedules to be reduced, the URAA has largely resolved the issue of the "exception" for primary products. And the constraints on domestic support have had the effect of restricting the ability of countries to reproduce by domestic subsidies the protection levels previously granted by reduced tariffs and export subsidies. Thus the rule changes have to a large extent met the need to incorporate agricultural trade in a rules-based trade system.

The impact that the URAA has had on individual countries varies greatly. All countries converted nontariff barriers to tariffs and bound those tariffs with the sole exception of rice quotas in Japan and Korea, which were allowed as temporary exceptions. Developing countries were allowed to declare "ceiling bindings" in place of product-by-product calculations of tariff equivalents, however. These ceiling bindings were commonly set at levels up to 100 percent or more, and thus had little impact on the actual level of tariffs used and the degree of market access. Tariffication had more impact in developed countries, where the quantitative restrictions were usually associated with sensitive products. In these cases the degree of market opening depended on the size of the TRQ agreed upon and the administration of that quota. Many countries considered the increased trade generated by the market access provisions of the URAA disappointing, and this increased the pressure for substantial market opening in the Doha Round.

The constraints on export subsidies have generally been successful, in that countries have appeared to stay within their scheduled limits for those subsidies included in their schedules. WTO panels have found (notably in the Canada dairy, U.S. cotton, and EU sugar cases), however, that there have been subsidies that were not included in the schedules, and the panels have declared these to be prohibited. Domestic subsidy constraints have also been generally respected, mainly because domestic policies in developed countries have tended to switch away from Amber Box subsidies. But there is continued concern that such subsidies cause considerable harm to other countries, and this has been confirmed by the panel in the U.S. cotton case.

Agreeing on disciplines on agricultural trade (as well as ending the quota system for textile imports) was a major step in completing the agenda embodied in the GATT of bringing all sectors in goods trade under the same regime. All agricultural tariffs are now bound, though they remain at a level several times higher than for manufactured goods. Nontariff barriers are no longer used, though TRQs still restrict market access. Though it does not directly mandate the type of policy instruments countries can use, the URAA has in effect provided a template for domestic policymakers: if they use WTO-compatible policies for their farm sectors they will be free from the constraints of the URAA. Export subsidies are still used but in much more restricted ways. The agricultural talks in the DDA have attempted to build on the achievements of the URAA.

See also agricultural trade negotiations; agriculture; Doha Round; multilateral trade negotiations; tariff rate quotas; Uruguay Round; World Trade Organization

FURTHER READING

Josling, Timothy E., Stefan Tangermann, and Thorald K. Warley. 1996. *Agriculture in the GATT.* Basingstoke, UK: Macmillan. A fuller account of the treatment of agriculture in the GATT.

Organisation for Economic Co operation and Development (OECD). 2001. *The Uruguay Round Agreement on Agriculture: An Evaluation of its Implementation in OECD Countries.* Paris: OECD. A useful study on the implementation of the URAA.

World Trade Organization. 1996. *The Results of the Uruguay Round of Multilateral Trade Negotiations: The Legal Texts.* Geneva: WTO. The full text of the URAA and other agreements.

——. 2000a. "Domestic Support." Background Paper by the Secretariat. G/AG/NG/S/1 (13 April). Geneva: WTO. A useful paper on the compliance of countries with domestic support commitments.

——. 2000b. "Export Subsidies." Background Paper by the Secretariat. G/AG/NG/S/5 (11 May). Geneva: WTO. A useful paper on the compliance of countries with export subsidy commitments.

TIM JOSLING

■ Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS)

The Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) has become the most important and far-reaching international accord in the field of intellectual property. It establishes workable global standards of protection and enforcement for virtually all of the most important intellectual property rights, such as patents, copyrights and related rights, and trademarks, in a single agreement. As such, it has major implications for knowledge-based industries seeking to trade profitably in many different countries.

History Strictly speaking, TRIPS is annex 1C of the Agreement Establishing the World Trade Organization, which was the main outcome of the Uruguay Round trade negotiations held under the auspices of the General Agreement on Tariffs and Trade (GATT). It resulted from a considerable amount of lobbying by certain industries that were keen to expand their activities in emerging economies where intellectual property protection was either lacking or was weakly enforced.

The first attempt to frame intellectual property as an issue to be discussed in wider trade negotiations was made by a group of trademark-holding firms organized as the Anticounterfeiting Coalition, which lobbied for the inclusion of an anticounterfeiting code in the 1973 79 GATT Tokyo Round. Although this initial attempt was unsuccessful, the copyright, patent, and semiconductor industries decided during the early 1980s to frame the lack of effective intellectual property rights protection in overseas markets as a trade-related issue *and* a problem for the U.S. economy that the government ought to respond to. By the time the contracting parties of the GATT met in Punta del Este, Uruguay, in September 1986 to launch another trade round, U.S. corporations had forged a broad cross-sectoral alliance and developed a coordinated strategy.

For those seeking high standards of intellectual property protection and enforcement throughout the world by way of the GATT, the strategy had three advantages. First, if successful, the strategy would globalize these standards much more rapidly than could be achieved through the conventions administered by the World Intellectual Property Organization (WIPO). This is because it allowed for the possibility of including all the main rights in a single agreement, which could also incorporate by reference provisions of the major WIPO conventions. Also, once it was agreed that the Uruguay Round agreements had to be accepted as a package (i.e., a "single undertaking"), countries could not opt out of any one of them and be a member of the new World Trade Organization (WTO). Second, the GATT already had a dispute settlement mechanism. WIPO has no enforcement or dispute settlement mechanisms except through the treaties that it administers, and these treaties do not provide much recourse for countries concerned about the noncompliance of other parties. Third, the broad agenda of the Uruguay Round provided opportunities for linkage-bargain diplomacy that WIPO, with its exclusive focus on intellectual property rights, did not allow. Hard bargaining by the United States, Europe, and Japan on intellectual property could thus be linked to concessions in such areas as textiles and agriculture, where exporting countries in the developing world were eager to achieve favorable settlements.

The Punta del Este Declaration of September 1986 included "trade-related aspects of intellectual property rights, including trade in counterfeit goods" as a subject for negotiations in the forthcoming trade round, which became known as the Uruguay Round. In full, the declaration's provisions on intellectual property were as follows:

> In order to reduce the distortions and impediments to international trade, and taking into account the need to promote effective and adequate protection of intellectual property rights, and to ensure that measures and procedures to enforce intellectual property rights do not themselves become barriers to legitimate trade, the negotiations shall aim to clarify GATT provisions and elaborate as appropriate new rules and disciplines.
>
> Negotiations shall aim to develop a multilateral framework of principles, rules and

disciplines dealing with international trade in counterfeit goods, taking into account work already underway in GATT.

These negotiations shall be without prejudice to other complementary initiatives that may be taken in the World Intellectual Property Organization and elsewhere to deal with these matters.

According to Susan K. Sell (2003), TRIPS is a case of 12 U.S. corporations making public law for the world. This makes sense only if one takes it to mean that the active engagement of these firms was a necessary, but not a sufficient, condition for there being a TRIPS Agreement. And actually she does not claim that the alignment of so much economic power and political influence made their victory inevitable or complete. Similarly, as John Braithwaite and Peter Drahos (2000) have noted, "It was a remarkable accomplishment to persuade 100 countries who were net importers of intellectual property rights to sign an Agreement to dramatically increase the cost of intellectual property imports."

So how was such a difficult feat achieved? Certain individuals played a decisive part in mobilizing support for the inclusion of trade-related intellectual property rights as a major Uruguay Round agenda item with the aim of formulating a legal instrument that would bind all members of what would become the WTO. These included chief executive officers of major corporations, lawyers, and a private consultant, all of whom were instrumental in conceptualizing intellectual property as a trade-related issue and then developing the political strategy that would ultimately result in TRIPS.

The interest groups succeeded in influencing the development of trade law and policy by incorporating their demands in the relevant legislation and by working closely with the key government agencies engaged in trade policy, especially the Office of the United States Trade Representative (USTR). Once the USTR had been persuaded that it was in the interests of the country to pursue the intellectual property demands coming from these groups, at least for the time being, at GATT rather than at WIPO, the next task was to form an international alliance

including the businesses and governments of Western Europe and Japan while neutralizing resistance from opposing countries.

Initially, the Group of Ten developing countries within the GATT India, Brazil, Argentina, Cuba, Egypt, Nicaragua, Nigeria, Peru, Tanzania, and Yugoslavia took a determined stand against the use of GATT as a forum for negotiating global intellectual property standards. But from 1985 and especially 1989 onward, the United States used its own trade rules to publicly criticize, threaten, and punish individual countries whose intellectual property standards were lower than its own and therefore "inadequate." Section 301 (Actions by U.S. Trade Representative) of the U.S. Trade Act was amended in 1984. The amended section 301 specifically included failure to protect intellectual property as one of the "unfair trade practices" that could result in a USTR investigation and possible sanctions, and authorized the USTR to initiate its own cases so as to protect U.S. firms from retaliatory action by foreign governments.

The 1988 Omnibus Trade and Competitiveness Act in its special 301 provision further strengthened the authority of the USTR in order to insulate decision-making on trade retaliation from foreign policy or national security considerations, and required the USTR annually to "identify those foreign countries that deny adequate and effective protection of intellectual property rights, or deny fair and equitable market access to United States persons that rely upon intellectual property protection." It is largely due to the mandate of the USTR to actively pursue the complaints of U.S. firms and business associations that the developing countries eventually accepted TRIPS.

Nonetheless, representatives of the United States, Europe, and Japan did not just sit down together and write the TRIPS Agreement themselves. Not only did divisions emerge between Europe and the United States that required compromises, but developing countries were much more involved in the drafting than they are often given credit for. As Jayashree Watal (2001) explains, they achieved favorable language in 10 of the 73 articles, albeit with the necessary

support of a few developed countries. The 10 include those dealing with the objectives and principles of TRIPS, limitations and exceptions to copyright, exceptions to patents and compulsory licensing, and control of anticompetitive practices in contractual licensing.

The Agreement's Key Provisions The preamble affirms the desire of member states "to take into account the need to promote effective and adequate protection of intellectual property rights," while "recognizing the underlying public policy objectives of national systems for the protection of intellectual property, including developmental and technological objectives." Dealing with counterfeiting is clearly considered as important. Its main importance lies in the fact that the trade in counterfeit goods is what makes intellectual property most clearly trade related. The preamble indicates that members recognize "the need for a multilateral framework of principles, rules and disciplines dealing with international trade in counterfeit goods." Yet the objectives as stated in Article 7 make no reference to the eradication of counterfeiting. Rather, TRIPS is explicitly aimed at promoting public policy objectives, the nature of such objectives presumably being left to national governments, though technological development is given priority.

Article 8.1 allows member states implementing their intellectual property laws and regulations to "adopt measures necessary to protect human health and nutrition, and to promote the public interest in sectors of vital importance to their socio-economic and technological development." These measures are not obligatory but, again, they highlight the socioeconomic welfare implications of intellectual property. On the other hand, the proviso that such measures be consistent with the provisions of TRIPS appears to narrow their possible scope quite considerably.

By virtue of Article 3, members accept the principle of national treatment, that is, that each country must treat nationals of other members at least as well as it treats its own nationals. In other words, intellectual property protection and enforcement must be nondiscriminatory as to the nationality of rights holders.

Article 4 upholds the principle of most-favored-nation. This means that any concession granted by one member to another must be accorded to all other members "immediately and unconditionally." So if country A agrees to take special measures to prevent the copying of the products of a company from country B, but turns a blind eye when the company is from country C, D, or E, such inconsistency of treatment will violate this principle. Although this principle of international law dates back in history, TRIPS is the first multilateral intellectual property treaty that refers to it.

Part II of TRIPS deals with the actual rights. These are very comprehensive, comprising the following:

1. Copyright and related rights
2. Trademarks
3. Geographical indications
4. Industrial designs
5. Patents
6. Layout-designs (topographies) of integrated circuits
7. Protection of undisclosed information
8. Control of anticompetitive practices in contractual licenses

To some extent the provisions are based on existing agreements. Thus WTO members are required to implement substantial parts of the Paris Convention on the Protection of Industrial Property and the Berne Convention of Literary and Artistic Works whether or not they are signatories to them. Nonetheless, while most developed countries were required only to make cosmetic changes to their intellectual property laws, most developing countries needed to reform their laws quite drastically. This is not surprising since the intellectual property standards provided in TRIPS tend to be modeled on the laws of the United States, Europe, or are a hybrid of the rules of the two jurisdictions.

All countries had to apply Article 3, on national treatment and on most-favored-nation status, and Article 5, concerning multilateral agreements on acquisition or maintenance of protection, within one

year of the entry into force of the WTO Agreement. But the developing countries and the former centrally planned socialist states were allowed a period of five years from the date of entry into force of the WTO Agreement, that is, until January 1, 2000, to apply the full provisions of TRIPS. Developing country members that were required to extend patent protection to areas of technology not hitherto covered in their laws were permitted to delay such extension until January 1, 2005. The least-developed countries (LDCs) were allowed until January 1, 2006, to apply TRIPS in full. Countries that have joined the WTO since then are required also to comply with these deadlines.

However, the LDCs have managed to secure two extensions. The 2001 Doha Declaration on the TRIPS Agreement and Public Health allowed them to delay implementation of patent protection for pharmaceutical products and legal protection of undisclosed test data submitted as a condition of approving the marketing of pharmaceuticals until January 1, 2016. In November 2005, the TRIPS Council extended the deadline to LDCs for fully implementing the rest of TRIPS by a further seven and a half years to July 1, 2013.

TRIPS places much emphasis on enforcement. With respect to the general enforcement obligations, procedures must be fair, equitable, and not unnecessarily complicated, costly, or time consuming. The judicial authorities must be granted the power to require infringers to pay damages adequate to compensate the right holder for the injury suffered due to the infringement. Members are required to provide for criminal procedures and penalties "at least in cases of willful trademark counterfeiting or copyright piracy on a commercial scale."

The agreement sets out the role of the Council for Trade-Related Aspects of Intellectual Property Rights (TRIPS Council). Accordingly, the council is responsible for:

   Monitoring the operation of TRIPS, and in particular members' compliance;
   Affording members the opportunity to consult on matters relating to trade-related intellectual property rights;

   Assisting members in the context of dispute settlement procedures; and
   Carrying out other duties assigned to it by the members.

The council is supposed to review the implementation of TRIPS at two-year intervals from January 2000. Article 71.1 states in addition that "the Council may also undertake reviews in the light of any relevant new developments which might warrant modification or amendment of this Agreement."

TRIPS-Related Developments at the WTO TRIPS was, and continues to be, highly controversial. Indeed, for both developing and developed countries, it represents unfinished business. Developing country representatives continue to express concerns that TRIPS raises prices of drugs and educational materials in poor countries, legitimizes the "biopiracy" of genetic resources and traditional knowledge, and blocks transfers of much-needed technologies. They have successfully resisted the further tightening of TRIPS rules and have had some small victories along the way. They have enhanced their capacity to put forward substantial counterproposals relating to such matters as public health, least-developed countries, traditional knowledge, and the compatibility between TRIPS and the Convention on Biological Diversity's provisions concerning benefit sharing, protection of traditional knowledge, and technology transfer. As for the developed countries and international business, which are constantly seeking ever higher levels of intellectual property protection and enforcement, TRIPS has to some extent been a disappointment.

It is in fact far from clear that making the intellectual property rules more or less identical whether you are a very rich country with enormous balance of payments surpluses in intellectual property protected goods, services, and technologies, or a poor country with highly burdensome trade deficits, is beneficial for the latter type of nation. While it is impossible to reliably calculate the long-term economic impacts of TRIPS on developing countries and their populations, we can be certain that they will incur short-term costs in such forms as rent transfers and administration and enforcement outlays, and that these

will outweigh the initial benefits. The cost-benefit balance will vary widely from one country to another, but in many cases the costs will be extremely burdensome.

At the November 2001 Doha Ministerial Conference of the WTO, members agreed on the texts of two very significant documents with provisions concerning intellectual property: the Ministerial Declaration, and the Declaration on the TRIPS Agreement on Public Health. The former declaration's TRIPS-related matters concerned geographical indications, the relationship between TRIPS and the Convention on Biological Diversity (CBD) and the protection of traditional knowledge and folklore, and technology transfer. The latter declaration dealt exclusively with TRIPS, primarily its public health provisions relating to compulsory licensing and parallel importation.

TRIPS Article 27.3(b) concerns exceptions to patentability in the area of biotechnology. It permits WTO members to exclude from patentability "plants and animals other than micro-organisms, and essentially biological processes for the production of plants or animals other than non-biological and microbiological processes." At Doha, ministers representing WTO members clarified their commitment to opening up negotiations on issues relating to Article 27.3(b) to include the relationship between the TRIPS Agreement and the CBD, and the protection of traditional knowledge and folklore.

The key challenge for developing countries is that many of them remain unclear about how to tailor their patent regulations to promote their interests in the acquisition, development, and application of biotechnology, and therefore how best to exploit the flexible language of Article 27.3(b). Understandably, though, much of the discussion has focused not specifically on this issue, but on how best to address a wide range of moral, political, and economic concerns about "patenting life" and "biopiracy."

The CBD-TRIPS relationship and the protection of traditional knowledge and folklore have proved to be quite controversial. One key developing country demand that has been pushed quite strongly is that of disclosure of origin. Disclosure of origin would re-quire inventors to disclose the source of genetic resources and/or traditional knowledge relevant to an invention being patented. In May 2006, Brazil, India, Pakistan, Peru, Thailand, and Tanzania proposed in the WTO General Council that new text be incorporated into the TRIPS Agreement under Article 29, which deals with conditions on patent applicants, to require such disclosure.

Geographical indications (GIs) are defined in the TRIPS Agreement as "indications which identify a good as originating in the territory of a Member, or a region or locality in that territory, where a given quality, reputation, or other characteristic of the good is essentially attributable to its geographical origin." In November 2001, the WTO members attending the Doha Ministerial Conference agreed "to negotiate the establishment of a multilateral system of notification and registration of geographical indications for wines and spirits by the Fifth Session of the Ministerial Conference." With respect to the possible extension of the enhanced protection of geographical indications to products other than wines and spirits, it was agreed that issues related to this matter would be addressed in the Council for TRIPS, an indication of the lack of consensus.

Despite the fact they are in TRIPS largely at the instigation of the European Commission, GIs have for several years been promoted as a concession to developing countries that they ought to take advantage of. Supposedly, they provide the means by which developing countries can use intellectual property to protect categories of local rural knowledge that they possess in abundance. In particular, the European Union and the Swiss government are very keen to promote GIs worldwide by arguing that this part of TRIPS can potentially provide substantial gains for developing countries. This seems plausible when one considers that GIs are especially appropriate for the produce of small-scale producers and cultivators, and, it should be underlined here, not just for foods and beverages but also handicrafts and other hand-made items.

Many developing countries are rich in traditional knowledge having applications in agriculture, food

production, and small-scale manufacturing. So GIs would appear to have real potential in terms of developing and exploiting lucrative markets for natural product based goods, including those manufactured by resource-poor farming communities. Such countries tend to favor the extension of the additional protection to cover all products, not just beverages. Are they right to be so pro-GI with respect to products they wish to export? Possibly they are, but caution should be exercised. GIs are useless without good standards of quality control and marketing, and up-to-date information on markets including foreign ones if the products are to be exported. At present the potential of geographical indications for developing countries is somewhat speculative because this type of intellectual property right has been used only in a few countries outside Europe. Moreover, many GIs have quite small markets, and a relatively small number are traded internationally.

Other developing countries do not have an abundance of traditional knowledge and are key exporters of products that compete with well-established GI-protected goods coming from Europe. For those countries, GIs may be more of a threat than an opportunity.

Multilateralism, Bilateralism, and the Future of TRIPS Developing country WTO members have been very reluctant to engage in negotiations to raise levels of intellectual property protection at the WTO. In order to hold these countries to more rigid and higher standards of intellectual property protection than TRIPS compliance requires, the United States and the European Union have gone outside the multilateral WTO forum. One of the most effective strategies being employed is that of bilateral and regional free trade agreements, which generally contain so-called "TRIPS plus" intellectual property right provisions, which place obligations on governments to provide more extensive protection than TRIPS actually requires. A growing number of developing countries seeking to enhance access to developed world markets for goods produced in their nations have proved willing to overcome their reservations about strengthened intellectual property rights through such deals in order to achieve this.

As a WTO agreement, and one which deals with arguably the most valuable assets of modern corporations, intangible ones, the importance of TRIPS to the world economy is immense and will remain so for several years. However, there are early signs that it is outliving its purpose for those corporations that successfully lobbied for an intellectual property agreement in the Uruguay Round and the governments that took up their demands. There are three reasons for this. First, the WTO system of trade governance currently does not make it easy to achieve radical revision of existing agreements or, for that matter, consensus on the need for new ones. Second, developing countries have tended not to implement TRIPS with much enthusiasm, and enforcement measures continue to be inadequate from the view of the intellectual property owners. Third, for the developed countries and transnational industry, other forms of trade diplomacy including those presented above seem to further their interests more effectively. Thus one may reasonably question whether TRIPS in the coming years will continue to be such an important agreement as it is today.

See also access to medicines; intellectual property rights; nondiscrimination; parallel imports; World Intellectual Property Organization; World Trade Organization

FURTHER READINGS

Braithwaite, J., and P. Drahos. 2000. *Global Business Regulation*. Cambridge: Cambridge University Press. Contains an excellent chapter explaining the political background to the incorporation of intellectual property in the Uruguay Round and the drafting of the resulting agreement.

Matthews, Duncan. 2002. *Globalizing Intellectual Property Rights: The TRIPS Agreement*. London: Routledge. Useful empirical study on various aspects of TRIPS from the viewpoint of European and global business interests and the developing countries.

Ryan, Michael P. 1998. *Knowledge Diplomacy: Global Competition and the Politics of Intellectual Property*. Washington, DC: Brookings Institution Press. The first book length account of how powerful political and business actors, based mainly in the United States,

organized a successful campaign that resulted in TRIPS.

Sell, Susan K. 2003. *Private Power, Public Law: The Glob alization of Intellectual Property Rights*. Cambridge: Cambridge University Press. Empirical and theoretical study that explains why the developed countries decided to push for a comprehensive intellectual property agreement in the Uruguay Round and not at WIPO. It also considers some of the economic and social welfare implications of TRIPS.

UN Conference for Trade and Development and Interna tional Center for Trade and Sustainable Development. 2005. *Resource Book on TRIPS and Development*. Cam bridge: Cambridge University Press. Definitive guide to TRIPS from a development perspective, which covers the negotiating history of the various articles, provides a technical interpretation of their language, and analyzes the WTO disputes relating to TRIPS. The chapters are separately downloadable from http://www.iprsonline .org/unctadictsd/ResourceBookIndex.htm

Watal, Jayashree. 2001. *Intellectual Property Rights in the WTO and Developing Countries*. New Delhi: Oxford University Press. Excellent insider account of TRIPS from a former Indian government trade negotiator and present WTO staff member.

World Bank. 2001. *Global Economic Prospects and the De veloping Countries 2002: Making Trade Work for the World's Poor*. Washington DC: World Bank. Chapter 5 provides a detailed study of the economic costs and benefits of intellectual property rights and the TRIPS Agreement.

GRAHAM DUTFIELD

■ **agricultural trade negotiations**

Opening up markets for agricultural products has proved a stumbling block for trade negotiations at both the regional and the multilateral levels. The primary reason is the political sensitivity of more open markets for farm and food products. Most governments share a concern for the security of their countries' food supply and the income level and stability of their rural sectors. In importing countries, this concern has led to caution about relying on imports for basic foodstuffs and a conviction that protection from overseas competition is necessary for the health of the rural economy. Those countries with export potential have long decried such sentiments, arguing that they can provide a regular supply of foodstuffs at lower prices and that supporting inefficient domestic production is not a sound basis for development. But, as one might expect in a sector where governments still have considerable control over markets, negotiations to open up trade in farm products have tended to proceed at the pace of the most reluctant importers.

Since the 1980s, this cautious attitude toward trade in farm goods has begun to give way to a more confident approach that sees imports as complementary to domestic production and exports as a natural extension of domestic markets. Consumers are becoming used to the greater choice of foodstuffs that comes with trade, and producers are setting their sights increasingly on foreign markets for new sources of revenue. In the process, many countries have become both importers and exporters of farm products and foodstuffs. This has blurred the easy categorization of a country's trade policy by its trade balance. Developing country importers often join with developing country exporters in voicing concerns about trade issues, particularly about the subsidies given to domestic producers in rich countries. Developed country exporters have "sensitive" sectors that apparently need to be sheltered even while they advocate more open markets for other products. Developed country importers with high protection barriers are often major importers of farm products needed for processing or for animal feed. Perhaps only in the market for tropical agricultural products is it still possible to identify typical "importer" and "exporter" views, but even in this case there are clear distinctions between those that have preferential access to markets in industrialized countries and those that do not benefit from such preferences, and between those that sell the raw materials and those that successfully add value in the domestic economy.

Along with this shift in political perceptions on agricultural trade has come a change in the nature of trade in farm products. In the 1980s, much of

the trade in primary agricultural products passed through sales or purchasing agents for producers and wholesalers, or of companies whose function was to distribute temperate-zone and tropical products through established channels. The role of state trading has shrunk markedly, with the adoption of policies to allow more private activity in marketing. Large companies now have a considerable role in the processing and marketing of farm products, as they have had for some time in the trading function. Most of these private actors operate in several countries, and thus food trade has become much more of a global business. The share of such trade that is categorized as "high value added" has correspondingly increased, leaving the trading of commodities and raw materials a smaller part of agricultural trade. Such trends explain the growing interest by large food and retail firms in removing trade barriers that act to inhibit worldwide marketing. As a result of these changes, in both political perception and structural reality, agricultural trade negotiations have been somewhat more successful in recent years in opening markets and have even made some progress in the past decade in reducing trade-distorting subsidies.

The change in the attitudes toward trade negotiations in agriculture is closely tied to reform of domestic policies. In developed countries, such policies were usually built on (and have been facilitated by) tariff and nontariff protection at the border and have employed a wide range of instruments to manage domestic markets and supplement farm incomes. Thus trade reform and market liberalization could not proceed as fast as in manufactured products in the postwar period. Domestic policies would have been impossible to maintain if trade liberalization had extended to agriculture. Reform of the domestic farm policies in developed countries started in the mid 1980s and continued apace for more than a decade. Policies that lowered support prices and substituted direct payments to farmers were found to be more easily amended to meet new targets, such as environmental stewardship, and tended to lower the incentive to produce unwanted surpluses. Developing countries, for different reasons, also relaxed their control over domestic markets and lowered trade barriers. In this case the motive was to correct macroeconomic and structural problems that were inhibiting development.

These reforms allowed countries to institute, as part of the Uruguay Round in 1994, a wide-ranging Agreement on Agriculture, which acted as a framework in which domestic policies could operate. This framework was consistent with lower protection at the border and less trade-distorting subsidies at home. It helped to lock in domestic reforms and put pressure on countries that were lagging in the reform process. But it also simplified the task of negotiating reductions in trade barriers and importantly changed the dynamic of such negotiations.

The same reform of developed-country farm policies also made it easier to negotiate bilateral and regional trade pacts. For years, most of these agreements had avoided the problem of negotiating reductions in tariff barriers for agriculture by explicitly excluding sensitive agricultural sectors from the full impact of market opening. This became insupportable when agricultural export interests, even in net importing countries, began to ask for market access (and preferences) within the regional or bilateral agreement. In addition, the rules of the World Trade Organization (WTO) (Article XXIV of the General Agreement on Tariffs and Trade 1994) oblige countries to grant tariff-free access on "substantially all trade" within bilateral and regional trade agreements. As a result, the inclusion of agriculture in such agreements is now the norm rather than the exception. Safeguards and slowly increasing tariff quotas still give some protection to sensitive domestic farm sectors, but few agreements exclude agriculture altogether.

It would be misleading, however, to suggest that these processes of globalization in farm and food trade and of reform of domestic policies have removed all the obstacles to open trade in agricultural goods. The process of reform has taken place at different speeds in different countries. So the pace of trade negotiations is still controlled to a large extent by the slowest reformers. Among the developed countries this includes Japan, Norway, and Swit-

zerland. Not only is protection high in these countries but the types of policies used still rely heavily on protection at the border. Hence they have been prominent members of the Group of 10 in the context of the WTO Doha Round, arguing for generous exclusions for "sensitive products," more modest tariff cuts, and no cap on the height of tariffs. Among the developing countries, the reluctant importers have formed the Group of 33, which emphasizes the need for adequate provision for "special products" and the inclusion of a "special safeguard mechanism" to allow them to reimpose tariffs if domestic markets are disrupted. These two groups, though negotiating actively, have effectively limited the "level of ambition" of the market access talks on agriculture.

Importantly, the process of trade and domestic reform has been uneven among sectors. The so-called white goods rice, cotton, sugar, and milk have among the highest tariff barriers and the most pervasive domestic subsidies. The sensitivity of these goods extends to exporting countries. The United States, normally a supporter of low tariffs, has come under pressure to reduce subsidies on each of these products. This has complicated the position of the United States in trade talks, arguing for others to open up markets but being more cautious in offering to cut support or lower tariffs where domestic interests are vocal. The European Union (EU) also has its sensitive products, including dairy and beef, though it has modified its domestic policies for rice, sugar, and cotton. The Uruguay Round did relatively little to improve the situation in the "white goods" markets, and their inclusion in regional trade agreements has often been politically sensitive. Sugar, for instance, was left out of the U.S.-Australia Free Trade Agreement, and was given a "temporary" exclusion from the Mercosur trade arrangements.

The implications of these changing political and economic forces can be seen with respect to the Doha Round of trade talks. Agriculture has been the biggest hurdle to an agreement, and the principal reason why talks were suspended for a time in July 2006. But the agricultural negotiations themselves are quite different from those in 1986, at the start of the Uruguay Round. At that time, the main protagonists were the United States and the EU, and the issues were whether to bring agricultural trade under the disciplines of multilateral rules and how to include rules for domestic policy that would be consistent with the trade disciplines. The United States, as a prominent exporter of temperate-zone farm products, favored such a move, in large part to circumscribe the Common Agricultural Policy of the EU. The EU had previously argued against this, on the grounds that domestic policy was a national issue and that the trade rules should give adequate scope for such internal choices. By the end of the talks, seven years later, the EU had agreed to rules that covered domestic policy instruments, categorizing them into "boxes" depending on their degree of trade distortion, and that eliminated quantitative trade barriers and limited export subsidies. The United States had achieved its objective of introducing binding rules but had to settle for only modest tariff reductions. The EU had to reform its own Common Agricultural Policy in 1992 in order to be able to live within the new constraints of the WTO.

The Doha Round is not about whether to develop rules for agricultural trade, or whether to extend trade rules to cover domestic policy. It was intended as a more "traditional" trade negotiation, to complement the outcome of the Uruguay Round by reducing tariffs by a substantial amount and by agreeing on a further reduction in subsidies. There is no longer any objection to the inclusion of domestic support in the trade talks and of curbing export subsidies. The focus has been on the depth of the tariff cuts and the extent to which certain "sensitive" products could be sheltered from the full cuts. In this sense it is much more in keeping with the negotiations in manufactured trade in the Kennedy Round (1963 67).

But one aspect of the Doha Round has made for more complex and difficult negotiations on agricultural trade. In contrast to the Uruguay Round, where with the exception of those in the Cairns Group few developing countries played a major role, the number of such countries that have been active has been remarkable. Spurred by the attempt (in August 2003) by the United States and the EU to develop a

common position, Brazil, India, and China, together with South Africa and a number of other developing countries, formed the Group of 20. The main demand from these countries was that the EU and the United States commit themselves to significant cuts in domestic support (subsidies) as well as eliminating export subsidies and cutting tariffs on farm goods. This coalition has stayed together and played a major role in the search for solutions to the agricultural negotiations.

Multilateral negotiations on agricultural trade have changed in nature as globalization has broken down the easy categorization of countries as reluctant importers and aggressive exporters. The scope for talks to open markets has been enhanced by the conversion of nontariff trade barriers to tariffs and the move to direct subsidies to support farm income. But the emergence of the major developing countries as players has moved the emphasis from transatlantic tensions to North-South conflicts. And so 21st-century trade negotiations, both in the WTO and in regional and bilateral trade pacts, have revolved around the extent to which developing countries need special rules to reflect their development status, how to continue the reduction of trade-distorting subsidies, and when to finally end the exception to normal trade rules that has allowed the continuation of export subsidies for primary products.

See also Agreement on Agriculture; agriculture; Doha Round; multilateral trade negotiations; tariff rate quotas; Uruguay Round; World Trade Organization

FURTHER READING

Coleman, William, Wyn Grant, and Tim Josling. 2004. *Agriculture in the New Global Economy*. Cheltenham, UK: Edward Elgar. Discusses the dimension of global ization that has begun to dominate agricultural trade from the viewpoint of political science.

International Center for Trade and Sustainable Develop ment (ICTSD). http://www.ictsd.org/about/index.htm. A source for current information on WTO agricultural negotiations and related issues, often from the viewpoint of developing countries.

Josling, Timothy E., Stefan Tangermann, and Thorald K. Warley. 1996. *Agriculture in the GATT*. Basingstoke, UK: Macmillan. The history of the place of agriculture in the General Agreement on Tariffs and Trade nego tiations prior to and including the Uruguay Round.

Sumner, Daniel A. and Stefan Tangermann. 2002. "Inter national Trade Policy and Negotiations." In *Handbook of Agricultural Economics*, vol. 2B, *Agricultural and Food Policy*, edited by Bruce L. Gardner and Gordon C. Rausser. Amsterdam, Netherlands: North Holland Press, pages 1999 2055. An overview of agriculture in trade negotiations.

World Trade Organization (WTO). http://www.wto.org/english/tratop e/agric e/negoti e.htm. The progress of negotiations in the Doha Round can be followed at this Web site.

TIM JOSLING

■ agriculture

Agriculture is the systematic raising of plants and animals for the purpose of producing food, feed, fiber, and other outputs. Historically, agricultural production has been linked to the use of the land and the tillage of the soil, and it is agriculture that allowed humans to initially establish permanent settlements. Such establishments were possible because agriculture provided the food necessary to meet human nutritional requirements. Because of its connection to the development of human societies, agriculture is closely associated with culture through the food that it produces and the manner in which it alters the landscape. This tie to culture gives agriculture special status in society and has led it to be treated differently from other commodities in international economic relations.

With economic development, the share of agriculture as a percentage of a country's gross domestic product (GDP) tends to decline (Chenery and Syrquin 1975). This trend is magnified on a global level. As countries have developed, the importance of agriculture in the global economy has declined, with agriculture in the early 21st century representing approximately 6 percent of global GDP.

This aggregate measure, however, masks the continued importance of agriculture for many

countries and households, particularly in developing countries. Because of its social, cultural, and economic importance, the treatment of agriculture in global economic relations has always been controversial. In Latin America, for example, approximately 20 percent of export earnings comes from agricultural products; in Africa it is 14 percent. Additionally, more than 40 percent of the world's economically active population still works in agriculture, with these workers mostly concentrated in developing countries (FAO 2005). Therefore, the operation of the global economy and its influence on the agricultural sector can influence the lives of billions of people.

Disputes over agricultural trade and domestic agricultural policies have continued to plague trade negotiations, including recent meetings of the World Trade Organization (WTO). These disputes have come at a time when agricultural markets have undergone significant changes, particularly in developing countries. A clear understanding of the role of agriculture in the global economy requires consideration of the importance of agriculture in global trade, its role in multilateral trade negotiations, its particular importance to developing countries, and recent trends that have transformed agricultural markets.

Agriculture in Global Trade The value of world agricultural trade nearly doubled between 1980 and 2000, from U.S. $243 billion to U.S. $467 billion. While average annual growth in agricultural exports was substantial during this period 4.9 percent in the 1980s and 3.4 percent in the 1990s it occurred at a time of generally increasing trade volumes and at a slower pace than growth in the manufacturing sector, particularly in the 1990s when manufacturing grew at an annual rate of 6.7 percent. This difference was even greater in developing countries, where agricultural export growth was 5.3 percent in the 1990s compared to 10.9 percent for manufacturing growth (Aksoy and Beghin 2005). Thus, while agricultural trade continues to expand and is clearly important to the world economy, there has been a general decline in its relative significance over time. For both the world in general and for

developing countries in particular, the share of agriculture in global trade has declined to just around 10 percent.

Along with a decline in the relative importance of agriculture, the composition of agricultural trade has shifted. First, there has been a movement away from the export of raw materials to greater export of processed products. Final agricultural products made up a quarter of world exports in 1980 81, but by 2000 2001 they had increased to 38 percent. Second, the commodities being produced and exported have changed from traditional tropical products (such as coffee, cocoa, tea, nuts, spices, fibers, and sugar) and temperate products (meats, milk, grains, feed, and edible oils) to nontraditional, higher-value products (seafood, fruits, and vegetables) and other products such as tobacco and cigarettes, beverages, and other processed foods. Tropical products, in particular, have declined from 22 percent of agricultural exports in 1980 81 to 12.7 percent in 2000 2001 (Aksoy and Beghin 2005). The expectation is that there will be a continued shift away from raw materials and traditional products toward these nontraditional higher valued products and processed items.

These overall trends in agricultural trade and volumes also mask the fact that much of the trade in agricultural commodities occurs within key trading blocs, particularly within the European Union (EU) and the member countries of the North American Free Trade Agreement (NAFTA). Table 1 shows the flows of global agricultural trade between different sets of countries in 2000 2001. Of the U.S. $181 billion in agricultural exports from the EU countries, U.S. $131 billion or 73 percent went to other EU member countries. Similarly, of the U.S. $90 billion in agricultural exports from the NAFTA countries, 39 percent occurred between Canada, Mexico, and the United States.

Agricultural policies account for part of the reason why agricultural trade stays within trading blocs. Historically, developing countries have often taxed agriculture, whereas developed countries have protected agriculture from outside competition and subsidized agricultural production. During the past two decades, changes in policies in many developing

Table 1
Global agricultural trade flows, 2000 2001 (US$ billion)

Importers	Exporters							
	Low income countries	Middle income countries	Developing countries	EU 15	Japan	NAFTA	Other industrial countries	Total imports
Low income countries	1.50	4.48	5.98	2.01	0.06	1.99	1.78	11.82
Middle income countries	9.20	48.44	57.64	22.85	1.74	23.42	10.71	116.36
Developing countries	10.70	52.92	63.63	24.86	1.80	25.41	12.49	128.18
EU 15	9.65	37.81	47.46	131.33	0.15	9.57	9.38	197.89
Japan	2.52	19.21	21.73	4.48		17.61	5.11	49.28
NAFTA	3.72	21.95	25.67	12.60	0.54	34.80	4.77	78.38
Other industrial countries	0.54	3.24	3.79	7.22	0.08	2.15	1.70	14.94
Total exports	27.14	135.13	162.27	180.84	2.57	89.55	33.45	468.67

Source: Aksoy and Beghin (2005) using COMTRADE data

countries such as the devaluation of overvalued exchange rates, the reduction of import restrictions on manufactured goods, and the elimination of agricultural export taxes have removed the anti-agricultural bias in these countries, leading to greater incentives to produce agricultural products. Developed countries, however, have continued to protect and promote agriculture through price supports and tariffs as well as direct subsidies. Within trading blocs, these protections are either not applied or are less restrictive and thus lead to greater agricultural trade within these regions.

The protection and subsidies to agriculture in developed countries influence their competitiveness in domestic and foreign markets. As discussed in the next section, this has become an important point of contention in trade negotiations. To calculate the degree of agricultural protection for farmers in these countries, the Organisation for Economic Co-operation and Development (OECD) has developed a method for estimating the total support to agricultural producers provided by a range of policies. The OECD's producer support estimate (PSE)

measures the monetary value of total gross transfers to agricultural producers arising from agricultural policies (OECD 2006). Compared with total gross farm receipts, this measure can be used to identify the share of farmer revenue that results from government policies. Table 2 provides the estimated percent PSE in farmers' revenue for selected OECD countries as well as for the OECD as a whole. The table also shows the share of support coming from market price supports, such as border tariffs, and the share coming from direct payments to farmers. The data indicate that with the exception of Australia, a significant share of farmers' revenue in developed countries is the result of agricultural policies in those countries. Furthermore, a substantial portion of the PSE is due to market price supports such as tariff and other forms of trade protection. However, there has been a decline in the PSE for each country in the table and the OECD as a whole between 1986 88 and 2003 5, and with the exception of Japan, there has been a shift away from market price supports toward direct subsidies. These agricultural policies, however, continue to influence the trade flows of agriculture into

Table 2
Producer support in selected OECD countries

	PSE as a percent of gross farm receipts		Market price support (share of PSE)		Direct subsidy (share of PSE)	
	1986 88	2003 05	1986 88	2003 05	1986 88	2003 05
Australia	8	5	50	0	50	100
Canada	36	22	52	46	48	54
European Union	41	34	87	50	13	50
Japan	64	58	90	91	10	9
United States	22	16	37	26	63	74
All OECD	37	30	77	57	23	43

Source: OECD (2006)

and out of developed countries. The sheer size of the economies where these policies are in place means they have a significant impact on global agricultural trade.

Agriculture and the World Trade Organization
The protection and support of agriculture in a number of developed countries has been justified on the grounds that agriculture deserves special consideration because it is linked to national culture. This special status has been recognized by the multilateral trade negotiations that have occurred under the General Agreement on Tariffs and Trade (GATT) and later under the WTO. Under the GATT, at the initial urging of the United States, certain agricultural sectors were exempted from the general prohibition in the agreement against quantity restrictions, as were export subsidies. As the EU took shape, it developed a Common Agricultural Policy (CAP) that provided general protection and support to farmers in all member countries and led to strong and increasing support by the EU for the special status of agriculture in multilateral trade negotiations.

The EU, the United States, and Japan remain strong supporters of the special status of agriculture. Countries that wish to export their agricultural products to these markets, however, argue that these policies are unfair since not only is access limited for certain commodities, but the subsidies provided to

producers in these protected markets give them an advantage over producers in countries without such subsidies. Developing countries tend to be particularly critical of agricultural supports since they are not in a budgetary position to support agriculture and believe agriculture is one area where they may have a comparative advantage.

When the WTO was created as part of the Marrakesh Agreement in 1994, a specific Agreement on Agriculture was designed to align agricultural trade rules with those of other products. The Agreement on Agriculture dealt with three general issues that influenced global agricultural trade: market access, domestic support, and export subsidies. For market access, the agreement required countries to convert quantity restrictions to tariffs. Once tariff equivalents of quantity restrictions were established, they were to be reduced. For domestic support, the agreement distinguished between trade distorting measures, which countries were required to reduce, and nontrade distorting measures. As part of the Agreement of Agriculture, developed countries agreed to reduce the total aggregate measure of support (AMS) by 20 percent by 2000, while developing countries agreed to reduce it by 13 percent by 2004. Nontrade distorting measures were exempt from reduction commitments. A final special category of domestic support measures was also exempt from reductions. These measures, which were used especially in the

EU, covered payments under programs designed to reduce the overproduction that occurred under other traditional market support payments. For export subsidies, no new subsidies could be introduced and existing subsidies had to be reduced by 36 percent in value by 2000 for developed countries and by 24 percent by 2004 for developing countries, although special conditions applied in some cases (Ingno and Nash 2004). The changes in the rules governing agriculture are reflected in the data presented in table 2, which shows a general reduction in producer support among OECD countries and a shift from market price support to direct subsidies.

The Agreement on Agriculture has generally been viewed as a step forward for trade negotiations; although it did not bring about substantial changes in the short run, it did set up a framework that aligned agriculture with other products. Its supporters hoped that setting up this framework would lead to further liberalization of agricultural markets in future rounds of WTO negotiations, particularly in the current Doha Round of trade negotiations. In fact, the Ministerial Declaration launching the Doha Round put agriculture as the first item on the agenda, indicating its priority in the negotiations. However, the Doha Round has failed to make substantial progress and the negotiations have faltered largely because of disputes over agricultural issues. For example, prior to a WTO meeting in Cancun in 2003 a group of developing countries, referred to as the G21, stated clearly the importance of dealing with agriculture issues such as protectionism and farmer subsidies. The Cancun meeting made no progress on these agricultural issues and was generally viewed as a failure. Following a WTO General Council meeting in July 2006 and the continued failure to make progress, the chairman of the WTO's Trade Negotiation Committee noted that agriculture is "key to unlocking the rest of the agenda" and that there was "no visible evidence of flexibilities" that would help solve existing problems (WTO 2006). Agriculture, thus, remains the principal stumbling block to further trade liberalization.

Complicating the disputes over protection and support of agriculture are additional issues that have emerged as the Doha Round of negotiations has progressed. First are market access barriers that have been put in place because of domestic concerns about methods used for production. While this is a broader concern for trade negotiations, it is particularly important for food products where health and safety concerns are significant and where countries may legitimately claim a right to protect their consumers. Although international agreements and standards exist, separating legitimate from illegitimate standards imposed by particular countries is not straightforward. The case of genetically modified organisms (GMOs) is a good example. Because of concerns over the health and environmental impacts of such products, the EU has enacted laws requiring the labeling and traceability of GMOs. However, the United States has challenged these policies as illegitimate restrictions on trade, arguing there is no evidence to support such concerns. The dispute remains unsettled (Zarrilli 2005). A second issue is the protection of geographical indications (GIs), names or labels used to indicate the geographic origin of certain products and, by extension, to indicate that these products have certain qualities or meet certain standards of production because they are from a certain region. Many food and beverage terms   such as Parmesan and Gorgonzola cheeses, Parma ham, Chianti, and Champagne   have come to be used generically although they have specific geographic origins. The EU and a number of developing countries argue that if the products are produced through a traditional, controlled manner in a specific region, their names should be protected and that existing protection is insufficient. The EU has proposed amending the current protection under the Trade-Related Aspects of Intellectual Property (TRIPS) agreement in favor of a mandatory multilateral system of registering these types of products (Evans and Blakeney 2006).

Food, Agriculture, and International Development Of the 1.2 billion poor people in the world, 75 percent live in rural areas (IFAD 2001). Therefore, although the contribution of agriculture to global GDP continues to decline in importance, it remains a fundamental component of the livelihood

strategies of these households and is critical to the advancement of the less-developed countries. Not only does agriculture provide food for the survival of the rural poor, it may be an important path out of poverty, particularly if the poor can take advantage of opportunities to produce nontraditional, high-value crops, which tend to be labor-intensive. The manner in which the global economy operates can alter the prices farmers fetch for their products, thus offering them incentives to produce more or less depending on the direction of price effects. Trade policies in developed countries therefore have an impact not only on developing economies, but potentially on the level of rural poverty in those countries.

Other than altering the incentives for farmers in developing countries, the policies of developed countries affect developing countries' agriculture through food aid. As a result of incentives for over-production in their agricultural policies, many developed countries have food surpluses. A portion of these surpluses is often used as food aid to developing countries. Although this seems a reasonable response to hunger, critics of food aid have argued that the provision of food creates disincentives for local production and policy. Increasing the supply of food commodities may cause prices of food to decline, thereby leading to lower incentives for farmers, rich and poor alike, to produce. The provision of food aid may also limit investment in agriculture by developing country governments that count on a regular supply of food aid to overcome limitations in food availability. This can limit the development of the rural economy and perpetuate a state of dependency. Most of these issues can be overcome with the judicious use of food aid in emergency situations and the careful management of aid programs when implemented, but the use of surpluses for food aid has the potential to cause significant problems for developing economies (Barrett and Maxwell 2005).

Given the importance of agriculture as a productive activity in developing countries, particularly in the poorest countries, the manner in which it is treated in global economic relations, and the agricultural policies enacted by key players in devel-

oped economies, can have a profound impact on economic development. Trade controversies are then closely linked to broader global issues of poverty and equity.

The New Agricultural Economy in Developing Countries In recent years, agricultural production in many developing countries has undergone profound changes, leading it to become increasingly oriented toward high-value global and urban product markets. These changes have mirrored many of the changes that occurred in developed countries earlier and have been driven largely by changes within developing countries. On the demand side, these include increased urbanization, the expanded entry of women into the workforce, and the development of a middle class all of which have led to greater demand for higher-quality and processed foods (Reardon et al. 2003). On the supply side, the opening of markets to foreign direct investment (FDI) influenced not only manufacturing sectors but also the domestic food market and the role of agro-industry in those markets. These domestic changes, combined with the expanded trade in processed and nontraditional products, have had a profound effect on the workings of the agricultural sector in many developing countries.

Referred to by some as the "new agricultural economy," these changes have led to new organizational and institutional arrangements within the food marketing chain as evidenced by new forms of contracts as well as by the imposition of private grades and standards for food quality and safety. Where open markets with a variety of sellers of different food commodities used to operate, there are now frequently supermarkets. Where small-scale processors of goods for the domestic consumer market such as potato chips used to exist, multinational corporations have taken over. Other multinationals have invested in the production of high-value crops such as fresh flowers and vegetables for immediate export to foreign markets. These supermarkets, processors, and exporters have specific requirements for the commodities they wish to purchase in terms of the quality of the goods and the timing of receipt. Products destined for the food market, in particular,

often must meet certain sanitary and phytosanitary standards. To ensure those requirements are met, these buyers often procure their goods through contracts with producers rather than through spot markets.

These rapid changes have brought new challenges and opportunities to farmers and policymakers in developing countries. These new markets provide significant opportunities because they tend to provide a price premium for quality and high-value products. But they also may be difficult to access and, in some cases, involve substantial risk. Often, new inputs and varieties are required and farmers must have sufficient information on grades and standards to be successful. Governments can play a critical role in facilitating access to information as well as providing assistance with the transition into these markets. This is particularly important for poorer farmers who often lack the resources to take advantage of new opportunities. Governments also play a critical role in verifying and certifying that sanitary and phytosanitary measures are in place. Furthermore, since such standards are often used as a mechanism to limit trade, governments may be called on to ensure that limits to market access are legitimate under the WTO rules.

The Role of Agriculture in the World Economy
While agriculture continues to decline in its overall economic importance in global trade, its special status in trade negotiations and its critical role in economic development and global poverty alleviation make it a sector that will continue to play a critical role in the world economy. In the short run, agricultural issues are likely to dominate trade negotiations. In the long run, global and domestic agricultural policies will influence the ability of countries to develop, and for the poor within those countries, to exit poverty. Agriculture will continue to be part of any discussion of global economic relations.

See also Agreement on Agriculture; agricultural trade negotiations; Common Agricultural Policy; distortions to agricultural incentives; primary products trade; sanitary and phytosanitary measures

FURTHER READING

Aksoy, M. Ataman, and John Beghin, eds. 2005. *Global Agricultural Trade and Developing Countries*. Washington, DC: World Bank. An edited volume that provides a broad overview of the issues related to global agricultural trade and developing countries.

Anderson, Kym, and Will Martin. 2005. "Agricultural Trade Reform and the Doha Development Agenda." World Bank Policy Research Working Paper 3607. Washington, DC: World Bank. Evaluates how the various regions of the world and the world as a whole could gain from agricultural trade reform.

Barrett, Christopher, and Daniel Maxwell. 2005. *Food Aid after Fifty Years: Recasting Its Role*. London: Routledge. An overview of the history of food aid, its impact on recipient countries, and its potential role.

Chenery, Hollis, and Moises Syrquin. 1975. *Patterns of Development, 1950 70*. London: Oxford University Press for the World Bank. Describes the structural changes that tend to occur over time as countries experience economic growth.

Evans, G. E., and Michael Blakeney. 2006. "The Protection of Geographical Indications after Doha: Quo Vadis?" *Journal of International Economic Law* 9 (3): 575 614. Discusses the possible steps forward for resolving the dispute over protection of geographical indications.

FAO. 2005. *The State of Food and Agriculture 2005: Agricultural Trade and Poverty, Can Trade Work for the Poor?* Rome: Food and Agriculture Organization of the United Nations. Assesses how agricultural trade can help reduce poverty and food insecurity.

IFAD. 2001. *Rural Poverty Report 2001: The Challenge of Ending Rural Poverty*. Oxford: Oxford University Press for International Fund for Agricultural Development. Examines rural poverty in the world and policies that may be used to reduce it.

Ingno, Merlinda, and John Nash. 2004. *Agriculture and the WTO: Creating a Trading System for Development*. Washington, DC: World Bank. An edited volume that examines the WTO agreement and the issues in the agreement facing developing countries.

OECD. 2006. *Agricultural Policies in OECD Countries: At a Glance 2006 Edition*. Paris: Organisation of Economic Co operation and Development. Estimates of support to agriculture for OECD countries.

Reardon, Thomas, C. Peter Timmer, Christopher B. Barrett, and Julio Berdegué. 2003. "The Rise of Super markets in Africa, Asia, and Latin America." *American Journal of Agricultural Economics* 85 (5): 1140 46. An overview of the transformation of the agri food system in Africa, Asia, and Latin America.

WTO. 2006. "Report by the Chairman of the Trade Ne gotiations Committee." http://www.wto.org/english/ news e/news06 e/tnc chair report 10oct06 e.htm. Report on the problems facing the Doha Round of WTO negotiations.

Zarrilli, Simonetta. 2005. *International Trade in GMOs and GM Products: National and Multilateral Legal Frame works. Policy Issues in International Trade and Com modities Study Series No. 29.* Geneva: United Nations Conference on Trade and Development. Discusses the national and international legal frameworks for man aging GMOs and the current issues being confronted in trade negotiations.

PAUL WINTERS

■ aid, bilateral

Bilateral aid is official development assistance that flows directly from a donor country government to a recipient country. This is in contrast to multilateral aid, in which many donor governments pool their contributions via intermediary institutions that then disburse aid to recipient countries. Yet another category is private aid that individuals, corporations, and foundations donate voluntarily, often through nongovernmental organizations (NGOs). Typically, bilateral aid accounts for two-thirds to three-quarters of all official aid; estimates of private aid put it at about one-tenth the size of official aid.

Although bilateral aid of one form or another is as old as the nation-state, the modern era of bilateral aid began with U.S. aid to reconstruct Europe under the Marshall Plan following World War II. Over time, governments of other major developed countries joined the United States in providing bilateral aid, sometimes following U.S. pressure to share the burden. In some cases, aid programs developed out of colonial administrations during the 1960s as more and more colonies became independent. In other cases, notably Japan, the bilateral aid program evolved from war reparation payments. A number of developing countries notably China, India, and Venezuela are also aid donors.

Bilateral aid is most often government-to-government, although in some circumstances donors may fund NGOs directly. Many aid donors are members of the Organisation for Economic Co-operation and Development (OECD). The Development Assistance Committee (DAC) of the OECD is the international body that sets aid reporting standards, monitors aid flows, and urges donors to improve the quality and quantity of their aid. The DAC's main focus is on official development assistance (ODA), which it defines as official concessional flows for developmental purposes to low-income countries. This includes grants as well as loans that are at least 25 percent concessional as compared to a commercial alternative (35 percent for "mixed credits," where aid is used to finance a commercial venture). Over the life of an ODA loan, repayments of principal and interest must be at least 25 percent lower than for a comparable commercial loan. ODA normally excludes grants and loans for military purposes and funds not directed to poor countries. In contrast to multilateral aid, most bilateral aid is given as grants; Japanese aid is the exception. The DAC's long-standing goal is for donors to contribute seven-tenths of 1 percent of gross domestic product (GDP) as ODA, though only a few of the most generous donors attain this target. In general, the United States has been the largest donor but among the least generous as a share of GDP. Under the Marshall Plan, U.S. aid was as high as 2 percent of GDP, but it has been one-tenth of that in recent times. France, Germany, and Japan also have been major donors, while the Netherlands and the Scandinavian countries have been the most generous DAC donors relative to their GDPs.

Bilateral aid can fund a specific project (project aid), provide more general budgetary support for the recipient government (program aid), or flow through an NGO. Although reconstruction aid to Europe was often program aid, bilateral aid to developing

countries has more often been project aid. Since the 1980s, multilateral aid agencies, particularly the World Bank, have become heavily involved in program aid, which aims to promote policy and institutional reform in developing countries; most bilateral donors put less funding into program aid.

Controversies Despite its apparent humanitarian nature, aid especially bilateral aid has been heavily criticized as insufficiently humanitarian and relatively ineffective. Critiques begin with the low volume of aid, both relative to the size of donor economies and relative to the need of the recipients. Critics also point to donor behavior that suggests that need can take a backseat to more narrowly defined donor interests such as geopolitics and commercial advantage. The geopolitical imperative is particularly pronounced for the United States with top recipients (Israel, Iraq, South Vietnam, South Korea) reflecting U.S. military interests rather than recipient need. For European donors former colony status dominates need, while Japanese aid often flows to countries rich in raw materials that Japan lacks.

The composition of aid also undercuts its development potential, as much bilateral aid is tied to purchases of donor products that are often expensive and inappropriate. Where aid funds projects, researchers have documented a bias toward large import- and capital-intensive undertakings that suit donors' needs but fail to reflect the relative scarcity of capital and foreign exchange in recipient countries. Aid levels are generally contingent on the donor's budget position so that aid tends to increase when the world economy is doing well. The result for many developing countries is procyclical aid fluctuations that can have the unintended consequence of destabilizing the recipient economy. The multiplicity of donors, each with their own priorities, procedures, and teams of visiting experts, can create a huge operational burden on a developing country government. All of these factors can reduce the development effectiveness of bilateral aid.

In response to the shortcomings of bilateral aid, the DAC and others have pushed for more multilateral aid. But if the main developmental short-comings of bilateral aid arise from its use as a geopolitical and commercial tool, why would donors be willing to redirect funds from bilateral to multilateral agencies? One argument is that the greater apparent independence of multilateral agencies makes them more efficient at some tasks (e.g., promoting sensitive institutional changes and providing a credible signal to private capital markets about the investment climate in the recipient country). A second argument is that donor country taxpayers favor developmental rather than geopolitical aid and that the donor government can demonstrate its developmental orientation most clearly through multilateral contributions. Neither argument, however, explains why some donors both contribute to and then work to undermine the independence of multilateral agencies. In any event, supporters of bilateral aid argue that the multiplicity of domestic interests served by bilateral aid helps to maintain a coalition in favor of larger budgets so that the net developmental impact of catering to domestic donor interests may be positive.

A recent trend in aid allocation has been toward selectivity. Research from the World Bank has argued that aid generally fails to promote growth and development when recipient government policies are poor and also that aid fails to promote policy change. Thus the implication is that more aid should be directed to countries that have already adopted appropriate (i.e., progrowth) policies. The empirical basis for these conclusions has proven weak. The original approach to estimating the link between aid and growth conditional on policy is not robust to small changes in specification or in which countries and years are examined. Other approaches have found a variety of results: aid has no impact regardless of policy; aid has a positive impact regardless of policy; aid has positive but diminishing returns regardless of policy; aid has a positive impact everywhere but the effect is larger when policies are good; and aid has a greater impact when policies are bad. Competing studies use different measures of aid, different approaches to deal with the potential for reverse causation, and different time horizons. The estimated impact of aid ranges from zero to eco-

nomically substantial (a several percentage point increase in the growth rate of GDP).

Despite this uncertainty, many bilateral donors have increased the country selectivity of their aid programs. In the United States, the Millennium Development Account is a direct application of aid selectivity with 16 indicators of good governance used as criteria for aid eligibility. Other donors, including Scandinavian countries, have moved toward using governance criteria to reduce the number of countries receiving funds. Although such a policy might seem to abandon poorly governed countries to their fate, advocates of selectivity maintain that incentive effects (the desire to qualify for aid) will induce better governance so that eventually all would benefit. In the past, however, political changes in donor governments have led to changes in bilateral aid allocation. If developing country governments expect such changes to continue, the incentive effects of current selectivity criteria will be undermined.

Opponents of aid have frequently rallied behind the slogan "trade not aid," pointing out that protectionist trade policies in donor countries cost developing countries far more than they receive in foreign aid. In addition, they argue that trade is likely to improve the efficiency of developing country economies whereas aid could introduce perverse incentives and promote rent seeking or corruption. Yet the arguments in favor of more trade do not weaken the case for aid as the slogan's "either/or" dichotomy seems to suggest. The distributional impact of trade is likely very different from that of aid, as the poorest in most need of aid are unlikely to be the main beneficiaries of increased trade.

See also aid, international; aid, international, and political economy; HIV/AIDS; nongovernmental organizations (NGOs); political economy of policy reform; regional development banks; World Bank

FURTHER READING

Burnside, Craig, and David Dollar. 2000. "Aid, Policies and Growth." *American Economic Review* 90 (4): 847 68. World Bank researchers present evidence in favor of greater aid selectivity; extremely influential even though the results were later shown to be fragile.

Lancaster, Carol. 2007. *Foreign Aid: Diplomacy, Development, Domestic Politics*. Chicago: University of Chicago Press. Provides an overview as well as a more detailed treatment of five major donors.

Minoiu, Camelia, and Sanjay Reddy. 2007. "Aid Does Matter, After All: Revisiting the Relationship between Aid and Growth." *Challenge* 50 (2): 39 58. Summarizes the literature on aid and growth and presents evidence that the long run impact of developmental aid on growth is large and positive regardless of the policy environment.

Organisation for Economic Co operation and Development. Yearly. *Development Cooperation Report*. Paris: OECD. Downloadable from http://www.oecd.org/dac. Information on aid issues, statistics, and practices.

Radelet, Steven. 2006. "A Primer on Foreign Aid." Center for Global Development Working Paper 92. Down loadable from http://www.cgdev.org/content/publications/detail/8846. Readable overview of foreign aid history and issues, including bilateral and multilateral aid.

Tendler, Judith. 1975. *Inside Foreign Aid*. Baltimore: Johns Hopkins University Press. Classic treatment of institutional biases reducing the effectiveness of aid. Uses evidence from U.S. bilateral aid programs.

CHRISTOPHER KILBY

■ aid, food

Food aid is a resource provided on concessionary terms in the form of, or for the provision of, food. In accounting categories, international food aid is a cross-border flow that is an entry into a country's balance of payments. National governments do provide food aid within their own borders, however. Generally, food aid is thought of as assistance provided by donor governments and humanitarian agencies to address the problems of hunger, food insufficiency, and malnutrition and indeed *some* food aid is expressly for this purpose. Over the history of modern food aid, however, substantial amounts of food aid have had little to do with addressing hunger.

History and Origins Food assistance throughout history has included attempts to address famine in

Venezuela in the 1830s, Ireland in the 1840s, and the Ukraine in the 1930s all of which involved significant shipments of food. The India Famine Codes in the late 19th century may have been the first systematic attempt to balance market forces and free food distribution. But the origin of large-scale contemporary food aid traces to the post World War II period in North America, when technological advances led to dramatic increases in production that resulted in significant surpluses and declining prices for farmers. Politically obligated to purchase the surpluses, the U.S. and Canadian governments found it less costly to give away the surplus food than to store it. United States Public Law 480, enacted in 1954, formalized this process. Hence the origin of modern food aid was a means of addressing domestic surplus production not a means of addressing hunger or poverty. Over the past 50 years food aid has served multiple objectives: disposing of surplus production, supporting farm income, maintaining a maritime shipping fleet, supporting domestic agribusiness and food processing industries as well as addressing hunger in foreign countries.

Types of Food Aid Classically there have been three kinds of food aid. *Program food aid*, for almost 40 years the largest single category, is subsidized deliveries or free grants of food on a government-to-government basis. The recipient government usually sells the food and uses the proceeds for many purposes not necessarily for food assistance or anything to do with addressing hunger. The main purpose of program food aid has been to provide budgetary support or balance-of-payments relief for recipient governments.

Project food aid provides support to field-based projects in areas of chronic need through deliveries of food, usually on a grant basis, to a recipient government, a nongovernmental organization, or the U.N. World Food Programme. The recipient agency then uses the food either directly in projects such as mother and child health, school feeding, or food-for-work projects that provide an employment guarantee using food for wages, or by "monetizing" the food aid selling it in the recipient country market and using the proceeds for project activities that require cash as an input rather than food.

Emergency or humanitarian food aid consists of deliveries of free food to populations affected by conflict or disaster, with a host country government, the World Food Programme, or a nongovernmental organization acting as the distributing agency.

Trends in Food Aid The total annual flow of food aid in the years 1981 through 2005 was around 8 10 million metric tons (see figure 1) a relatively tiny amount compared to the more than 300 million tons of commercially traded food of the same commodity groups. Since 1980, humanitarian food aid has become the most dominant form, amounting to about 60 percent of the total, whereas prior to 1980, it was in the range of 10 15 percent. Since the beginning of the 21st century, program food aid or government-to-government food assistance has declined sharply, and by 2007 seemed likely to be phased out altogether soon. Project food aid tended to be relatively stable in volume terms in the period 1980 2005. In the post World War II years, food aid comprised as much as 15 percent of total overseas assistance in the early 21st century, by contrast, it accounted for only 2 3 percent of the total.

Food Aid Policy Debates International food aid has been subject to weak regulatory mechanisms, many of which are outdated or dysfunctional. As a result, many of these mechanisms were being renegotiated as of 2007, or their roles were being taken over by other, newer institutions. The Food Aid Convention originally signed in 1967 and renewed
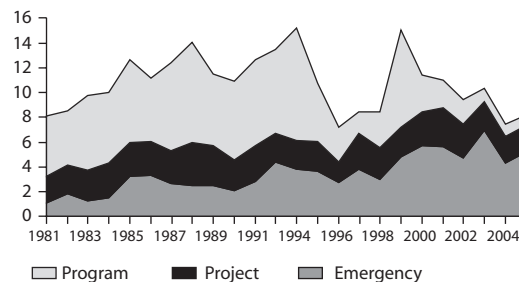
Figure 1

Global food aid deliveries, 1981 2005 (MMT). Source: WFP Interfais.

in 1998   existed on one-year extensions for several years in the first decade of the 21st century, and was up for a major renegotiation by 2007. Much of the impact of food aid on trade has come up for negotiation at the World Trade Organization.

The extent to which food aid undermines international trade or local incentives for agricultural production is also a subject of controversy. Overall, there appears to be no significant impact of food aid on domestic production, but different kinds of food aid have different effects   well-managed emergency food aid can actually lead to an increase in food production (with a lagged time effect, because well-managed emergency food aid can enable farmers to cultivate a crop when they might otherwise be unable, thus resulting in a greater harvest available at the end of the season). But poorly targeted food aid that ends up being consumed by people who would otherwise be able to grow or purchase food may undermine production or trade incentives. Open "monetization" (the sale of food aid in the recipient country) is thought to be one example that causes trade or production displacement.

Traditionally food aid was provided in kind by donor governments, to be used or distributed either by recipient country governments or by humanitarian agencies. Beginning in the 1990s, food aid from some donors has been "untied" from market sources in donor countries, meaning that a proportion of food aid budgets are now in cash for purchasing food either within the country where food aid is needed, or in a nearby developing country called local and regional purchase. European donors and Canada devote significant proportions of their funding to local or regional purchase of food, which is believed to save both time and expense when responding to acute need, thus enabling more rapid response and a relatively greater volume of response within fixed budgets. United States food aid continued to be tied to procurement in U.S. markets as of late 2007, and the extent to which "aid-tying" of food assistance continued to be allowed was a subject of considerable controversy in World Trade Organization negotiations over a new Agreement on Agriculture. In addition to pressure for untying

food assistance, there was also considerable evidence that cash transfers are a more effective and efficient means of ensuring access by poor or disaster-affected populations to adequate food under some circumstances.

See also Agreement on Agriculture; aid, international

FURTHER READING

Barrett, Christopher, and Daniel Maxwell. 2005. *Food Aid after Fifty Years: Recasting its Role*. London: Routledge. Traces the origins and history of food aid, current applications and problems, and policy proposals for addressing many contemporary problems with food aid.

Clay, Edward, and Olaf Stokke. 2000. *Food Aid and Human Security*. London: Frank Cass. An edited volume dealing with many of the issues related to food, including the right to food.

Harvey, Paul. 2007. *Cash based Responses in Emergencies*. HPG Paper. London: Humanitarian Policy Group. A state of the art review of using cash transfers in emergencies, to fulfill food security or other objectives.

The Sphere Project. 2004. *The Humanitarian Charter and Minimum Standards in Disaster Response*. Geneva: The Sphere Project. A humanitarian charter made up of key points from international human rights law, international humanitarian law, and refugee law. The "Sphere guidelines" also lay out minimum standards for assessment, service delivery, monitoring and evaluation, and participation in seven sectors of humanitarian intervention, including food aid and food security.

UN Food and Agriculture Organization (FAO). 2006. *The State of Food Insecurity in the World, 2006*. Rome: Food and Agriculture Organization. An annual report, the 2006 edition is devoted entirely to the subject of food aid.

UN World Food Programme. WFP Interfais. Available at http://www.wfp.org/interfais/index2.htm. Offers tables that detail recent allotment of food aid.

DANIEL MAXWELL

■ **aid, humanitarian**

Humanitarian aid is assistance provided directly to people affected by conflict or disasters. Traditionally it is in-kind material assistance   food, shelter, or

medical care   but it also can take the form of cash assistance. The primary intent of humanitarian aid is to save lives, reduce suffering, and protect human dignity in times of crisis. Often, additional objectives are to protect people's livelihoods and assets or other indirect life-saving goals. Humanitarian aid is different from development aid, which has longer-term objectives and sometimes greater conditionality but the distinction is not always clear, and in chronic emergencies may become quite blurred.

History and Origins Although there are humanitarian antecedents in many different religious and cultural traditions, the history of modern independent humanitarian aid traces to the Battle of Solferino in 1859. There, Swiss businessman Henri Dunant was horrified by the impact of industrialized warfare and organized a local effort to assist wounded soldiers on the battlefield. Dunant essentially struck a deal with the belligerent armies that he would not interfere with their conflict provided they would grant him safe passage to assist the wounded who were incapable of continuing to fight. In 1863, Dunant founded the International Committee of the Red Cross (ICRC) in Geneva, based on the same principle. The ICRC is the organization mandated by international humanitarian law to provide protection and assistance to wounded or captured soldiers, and especially to civilians and other noncombatants caught in conflict. National chapters of the Red Cross soon grew up as well, usually mandated by law in the host country. In Islamic countries, these are known as Red Crescent societies. The International Federation of Red Cross and Red Crescent Societies (IFRC) is made up of these national chapters.

Other Humanitarian Agencies Under international law, the state has the first responsibility for the protection and provision of assistance to citizens within its borders. But often the government of an affected country either cannot provide adequate assistance or, in some cases, may be one of the parties causing a humanitarian emergency in the first place. In such cases, other agencies may be called on (or take it on themselves) to provide assistance. Following each of the major crises of the 20th century, other agencies were founded to provide humanitarian

assistance for the victims of war as well as people affected by natural disasters or other crises. These include specialized UN agencies and nongovernmental organizations (NGOs). UN agencies mostly formed in the aftermath of World War II include the United Nations Children's Fund and the Office of the UN High Commissioner for Refugees. The largest of the UN humanitarian agencies, the World Food Programme was founded in 1963, and the Office for the Coordination of Humanitarian Affairs was founded in 1998, reorganized from the former Department of Humanitarian Affairs.

Major NGOs were also formed in response to the crises of the 20th century: Save the Children was formed in the aftermath of World War I, CARE and Oxfam in response to World War II, and World Vision in response to the Korean War. Médecins sans Frontières was formed in 1969 after French doctors split with the ICRC over the issue of raising public awareness about the plight of civilians during the Biafran war. More recent crises have seen the advent of increasing numbers of nontraditional providers of humanitarian aid   including the military, private for-profit companies, and contractors.

Defining Characteristics In principle, humanitarian assistance is based on the premise that no human caught in a life-threatening situation should be denied assistance (or conversely   and more controversially   that all humans in a life-threatening situation have a right to receive assistance). This was formalized by the ICRC in 1965 as the principle of *humanity* or the *humanitarian imperative*. This principle now informs humanitarian assistance across most if not all humanitarian agencies. The ICRC also formally articulated six other principles on which it had based its work for nearly a century and which followed logically. The first is *impartiality*, or the imperative to provide assistance on the basis of need alone, without respect to nationality, race, religion, gender, or political point of view. The second is *neutrality*, or a refusal to take sides in a conflict or to engage in "political controversies." This was seen as a principle in its own right and the means to gain access to conflict-affected populations. A third principle is *independence*, meaning that the offer of humanitarian

assistance should be independent of government policies or actions. *Voluntary service*, *unity*, and *universality* are the other principles. These principles are codified for the entire humanitarian community in the Red Cross Code of Conduct and the Sphere Guidelines.

Humanitarian Policy Debates Up to the 1980s, humanitarian action was seen to be largely beyond reproach — a form of altruism that was neither a threat nor of serious political consequence. After the end of the cold war, humanitarian action became, for a time, a substitute for major powers' political engagement in problems affecting other countries — for example, in the Balkans wars of the 1990s. Since the onset of the global war on terrorism, humanitarian action has come to be seen, both by major powers and, in some cases, by popular perception in host countries, as part of a larger political and security agenda. Most humanitarian agencies still refuse to take sides in conflicts, but neutrality — especially the admonition against "political controversy" — is the subject of intense debate and many humanitarian agencies do not claim to be apolitical (although the ICRC still does). And as public sources of funding for humanitarian aid become more and more dominant, the capacity for independent humanitarian action has declined.

The principle of impartiality is intended to govern the provision of humanitarian aid in all circumstances, but political considerations often trump humanitarian principles in practice, and there are large discrepancies in the level of response by the humanitarian community to similar levels of human suffering by affected populations — from as little as a few dollars per person in some forgotten crises in Africa to more than $7,000 per person affected by the Indian Ocean earthquake and tsunami emergency in late 2004.

With larger budgets and greater scrutiny have come calls for the humanitarian enterprise to be more accountable — not only to the donors that provide funding, but for the behavior of agency staff in emergencies, for the development of industry standards for humanitarian response to which all actors should be held responsible, and to the recipient communities that are affected by war and disaster. To this end, in the early 21st century several interagency initiatives, including the Sphere Project, the Action Learning Network for Accountability and Performance, and the Humanitarian Accountability Partnership International, developed to hold humanitarian action to higher standards of accountability.

See also aid, international

FURTHER READING

Darcy, James, and Charles Antoine Hofmann. 2003. *According to Need? Needs Assessment and Decision Making in the Humanitarian Sector.* HPG Report. London: ODI. A landmark report highlighting the fact that humanitarian response is not impartial (i.e., not according to the level of humanitarian need). The international humanitarian community has no means for independently assessing comparable levels of needs, and the donor community does not believe most humanitarian needs assessments.

Development Initiatives. 2006. *Global Humanitarian Assistance 2006.* Somerset, UK: Development Initiatives. A summary report on the level of funding, the types of emergencies encountered, and the types of agencies responding to humanitarian emergencies in 2006.

International Committee of the Red Cross. 1983. *Understanding Humanitarian Law: Basic Rules of the Geneva Conventions and their Additional Protocols.* Geneva: ICRC. A primer on international humanitarian law, for the "law of war," which governs all humanitarian action in times of war and spells out the obligations of all parties.

———. 1995. *Red Cross/NGO Code of Conduct.* Geneva: IFRC. The code of conduct that lays out the basic humanitarian principles for all agencies, whether affiliated with the Red Cross or not. The code of conduct goes beyond just the four main ICRC principles. The Code of Conduct has been signed by more than 400 agencies.

Rieff, David. 2002. *A Bed for the Night: Humanitarianism in Crisis.* New York: Simon and Schuster. A critique that is at once sympathetic to the humanitarian movement but harsh in its judgment of the loss of vision, direction, and principle in humanitarian action during the late 20th century. Published just after the attacks of September 11, 2001, it doesn't deal with humanitarian action in the context of the global war on terrorism.

The Sphere Project. 2004. *The Humanitarian Charter and Minimum Standards in Disaster Response*. Geneva: The Sphere Project. A humanitarian charter made up of key points from international human rights law, international humanitarian law, and refugee law. The "Sphere Guidelines" also lay out minimum standards for assessment, service delivery, monitoring and evaluation, and participation in seven sectors of humanitarian intervention.

DANIEL MAXWELL

■ **aid, international**

International aid, or official development assistance (ODA), comprises a wide range of financial and nonfinancial components. These may take the form of cash transfers as well as grants of machinery, technical advice, and analysis and assistance in capacity-building support. Although foreign aid is often envisaged as transferring resources from rich to poor countries, the reality is more complex, with more than half of all ODA actually going to middle-income countries.

Early Foreign Aid The history of modern aid and colonialism are in many ways intertwined. In order to extract raw materials and exploit economic activity abroad, the colonial powers provided investment capital, technology, and personnel to colonies. Examples include the Belgian-initiated railroads in the Congo, the French design of the Suez Canal, and railroads and roads built under British rule to transport primary commodities in southern and eastern Africa.

Explicit reference to aid became more widespread in the 1940s. During this period, the wealthier countries began considering broader economic development as a goal, focusing on aspects of engagement that were not directly or exclusively related to extraction and exploitation. This reflected an evolution of economic and strategic interests (and in part was associated with the decline of mining relative to other sectors and also with the development of air and other military capabilities). It also reflected a changing understanding of human dignity in which human rights and self-determination emerged in strong opposition to earlier Darwinian notions, which had reinforced colonial notions of superiority. The 1940s also brought to the fore arguments for the need to invest in strategic alliances and in peace.

The period immediately following World War II saw a concerted effort to avoid a repeat of the post World War I peace process and reparations, which had served at best as a short-term palliative. It was recognized that peace required economic integration and this required both financial flows and policy changes to bring about closer economic integration. The new vision for sustained peace through economic opportunity gave rise to the Marshall Plan, as well as to the Bretton Woods conference and the creation of multilateral institutions such as the United Nations (UN), the World Bank, and the International Monetary Fund, whose goal was to increase international cooperation and assistance. This architecture largely remains in place today.

Increasingly, broad goals such as education were supported by emergent foreign aid programs and endorsed in parliamentary acts of the time, such as the 1948 British Overseas Development Act. In 1949, President Truman's inaugural address proposed the creation of a program for development assistance. The UN's 1951 report on *Measures for the Economic Development of Under-developed Countries* (with Arthur Lewis as the lead author) advocated the creation of a dedicated UN fund to support development, as well as an International Finance Corporation (IFC) to underpin private investment (five years later the IFC was established as part of the World Bank Group). In the 1950s and early 1960s, following the granting of independence to most colonies, many former colonial civil servants were employed in newly established aid projects, so that while the expatriate personnel remained, the nature of the relationship changed.

The initial focus of these institutions was on rebuilding war-torn Western Europe and Japan and on stabilizing the world financial system, rather than on broader notions of development. Indeed, the first four of the World Bank's loans were devoted to postwar reconstruction. In addition to supporting

the activities of the multilateral agencies, the wealthiest countries began to set up their own bilateral initiatives to provide aid flows and technical assistance to developing countries. In 1960, Canada created an External Aid Office, and in 1961 the United States created the United States Agency for International Development, France inaugurated a Ministry for Cooperation, Japan created the Overseas Economic Cooperation Fund to provide loans for developing countries, and Germany established its development bank, the Reconstruction Credit Institute (KfW). Sweden and the United Kingdom established bilateral aid agencies in 1962 and 1964, respectively. These agencies at first focused on former colonies, leaving broader global reconstruction and development to the multilateral institutions such as the World Bank. Over time, however, the objectives and strategy of multilateral institutions and national agencies converged, and there is now a significant degree of overlap.

From the outset, nongovernmental and other groups had ambitions that were often ahead of governments' views on aid. For example, the World Council of Churches in 1958 called on the rich countries to allocate 1 percent of their national income to aid for developing countries. This was later taken up by the Development Assistance Group, which was established in 1960 as a forum for consultation among aid donors at the Organisation for European Economic Co-operation (OEEC), which became the Organisation for Economic Co-operation and Development (OECD) later that year. A target of 0.7 percent was agreed by the UN General Assembly in 1970, but by 2006 it had been reached by only five countries, despite the fact that rich countries are much wealthier and poor countries are much better managed than when the pledge was made.

Foreign Aid during the Cold War Chilling relations between the West and the Soviet Union meant that from the 1950s to the fall of the Berlin Wall in 1989, Cold War politics became a key determinant in all foreign policy, and not least aid. Increasingly aid was used as a means to support and bolster friendly states. Where geopolitical interests were involved, economic and military support were often closely interconnected.

In Zaire (now Democratic Republic of the Congo), for example, aid was used as a strategic tool. Between 1960 and 1990, more than $10 billion was disbursed in aid to Zaire in support of an increasingly brutal and corrupt dictatorship. Maintaining a strategic alliance, rather than development effectiveness, was the objective. Similarly, political and economic factors were the main drivers of large aid donations in support of the transition to a market economy in Eastern Europe and Central Asia. Not surprisingly, such aid was not correlated with long-term poverty reduction. Such cases have lent support to aid skeptics.

A key characteristic of foreign aid during the Cold War period was its "tied" nature: aid was allocated to the purchase of specific goods and services from the donor country. In this way, much aid never left the donor countries, as it ended up paying for consultants or services (such as foreign-language radio broadcasts).

Food aid also became prevalent. In part this reflected the growing agricultural subsidies in Western Europe, the United States, and Japan, which rose to more than $300 billion per annum. These rich countries protected their markets from competitive imports and subsidized their farmers, dumping surpluses on world markets, often as food aid. Poor countries were unable to compete in agriculture and trade, and the dumping of food on world markets undermined and destabilized agricultural prices and production in developing countries. As outlined by Goldin and Knudsen (1993) and Sen (1982), the distortion of markets and policy failures contributed to price instability, long-term decline in agricultural prices facing developing country farmers, and even famines. While protectionist trade policies in the rich countries undermined the potential for sustainable growth in many developing countries, rich countries' aid policies provided at best a partial response to short-term humanitarian needs. Similarly, while significant quantities of aid were directed into investments in irrigation and other agricultural infrastructure in developing countries, the undermining

of the rural economy as a result of protectionism by rich countries provides a classic example of a failure to achieve policy coherence.

Although there has been some progress in untying aid since the end of the Cold War, tying persists in many aid programs today. Meanwhile, agricultural protectionism continues to have a pernicious influence on world agricultural markets and trade, with the negative impact far exceeding the aggregate poverty-reduction impact of aid (Goldin and Reinert 2007).

Adjustment Programs Adjustment (or macroeconomic policy-based) aid was developed in the 1970s. This was designed to respond to the severe macroeconomic imbalances experienced by many poor countries, characterized by ballooning deficits and debt that were aggravated by exogenous shocks from oil prices, interest rates, and other sources. The constellation of conservative political leadership in the United States (Reagan), the United Kingdom (Thatcher), and Germany (Kohl) was also important, with aid acting as an agent of reform and bolstering the private sector. By the late 1970s, aid was increasingly predicated on the recipient country's acceptance of conditions that sought to enforce macroeconomic and trade reforms, and to facilitate private (particularly foreign) investment. The focus on structural economic reforms was not accompanied by corresponding attention to institutional reform or investments in education and health, and neglected demands of social cohesion. This limitation, along with the mounting assertions of self-determination by rapidly democratizing governments, as well as the pendulum swing in rich countries away from their former preoccupation with market solutions, meant that by the mid-1990s, following the end of the Cold War, adjustment lending had become far less common.

Poverty Reduction and Recent Development Models The reform of aid policies in the 1990s responded to a range of factors. These included the end of Cold War preoccupation with strategic allies, the success of the growth paradigms in China and India, the failure of major aid efforts (not least in Africa),

and intellectual evolution   particularly in the areas of understanding around economic growth, poverty reduction, and development. Perhaps the most striking belated change was the recognition that to reduce poverty, aid should be focused on those countries where poor people lived and in which the governments were willing and able to act to overcome poverty. In addition, as poverty increasingly was recognized as multifaceted, the policy discussions and interventions around poverty reduction became more nuanced.

By the late 1990s, the goals of development began to embrace the elimination of poverty in all its dimensions, by improving education, health, and other human capacities, not simply focusing on income. Scholars such as Sen (1999) gave intellectual form to the emerging understanding that development means increasing the control that poor people have over their lives. This is derived through a combination of education, health, and greater participation in politics and community decisions, as well as from improvements in access and income. It is also clear, and beginning to be reflected in aid policy, that the various dimensions of poverty are related, and that income growth generally leads to progress in the nonincome dimensions of poverty and vice versa.

From the early 1990s, this rethinking of development was associated with an increase in aid flows to health, education, and infrastructure. Due mainly to improvements in two countries, China and India, where government policies are largely but not entirely independent of aid, social indicators such as health and education on average have improved very quickly since the mid-1970s in developing countries.

Direct targeting of health and education goals, rather than waiting for improvements to follow income gains, has led to a virtuous circle that has improved the welfare of individuals and families. On average, at every level of income in developing countries, infant mortality fell sharply during the 20th century, and life expectancy increased by 20 years (from mid-forties to mid-sixties) over a period of only 40 years. The global trend in life expectancy remains very positive, although HIV/AIDS has dra-

matically reversed this trend in a number of countries, particularly in southern Africa.

Developing countries also experienced dramatic improvements in literacy: whereas in 1970 nearly two in every four adults were illiterate, now the proportion is only one in four. Particularly important is the greater focus on female literacy, as this has been shown to be pivotal in improving the health and human development levels of children and communities. These achievements cannot solely or even primarily be attributed to aid, but rather to a combination of national policies (not least in China and India) along with growing integration and the adoption of new health and other technologies.

The unevenness of the performance of developing countries, and ability of some very poor countries to grow despite resource constraints and others to squander abundant resources, has led to a growing focus of the aid community on governance and institutions. Increased attention to governance and institutions also has been associated with growing interest among academics in this topic. In the pragmatic world of aid agencies, attention to governance, and to institutional development more broadly, has been translated into growing budget allocations for the development of civil servants and building of regulatory authorities and judiciaries. It also has been associated with a growing emphasis on the allocation of aid in light of the performance of the recipients.

Millennium Development Goals In 2000 heads of state of both rich and poor countries committed themselves to achieving the Millennium Development Goals (MDGs). These are:

- eradicate extreme poverty and hunger
- achieve universal primary education
- promote gender equality and empower women
- reduce child mortality
- improve maternal health
- combat HIV/AIDS, malaria, and other diseases
- ensure environmental sustainability
- develop a global partnership for development

The Millennium Declaration marked a major step in the history of aid. For the first time, the international community came together to establish clearly defined common goals with a set of agreed measurable targets and results for developing countries, donor agencies, and the multilateral institutions. Building on this declaration, the 2002 Monterrey Conference established a new partnership for development in which the rich countries promised to increase both the volume and the quality of aid in return for commitments from developing countries to undertake vital reforms to enhance aid effectiveness. Despite the remarkable achievements in many countries, not least in China, it is becoming clear, however, that many or most of these goals will be missed, except perhaps that of halving income poverty, which will be met at the aggregate global level. Many developing countries have failed to live up to their Monterrey commitments on poverty reduction and good governance. Similarly, the rich countries have fallen well short of their commitments. Aid flows have increased, but these remain far short of the agreed targets, and the Doha Development Round of trade negotiations appears to be a long way from meeting even minimal expectations. The tremendous achievements of the MDGs, not least in terms of mobilizing public opinion in many rich countries, are therefore at risk of dissipating due to inadequate political will.

Types of Aid and Harmonization As figure 1 shows, only around 20 percent of bilateral aid in fact ends up as a cash transfer into the hands of the recipient country ("other bilateral"). Around 80 percent takes the form of aid to multilateral organizations, debt relief for countries such as Afghanistan and Iraq, administrative costs, costs for refugees living in donor countries, and technical cooperation including support to students from developing countries studying in donor countries (imputed student costs). Although this indirect assistance may make an important contribution, it is often driven by the priorities of the donors and is no substitute for predictable, multiyear flows of aid mobilized behind government programs that are agreed upon by governments and across the donor community. Such
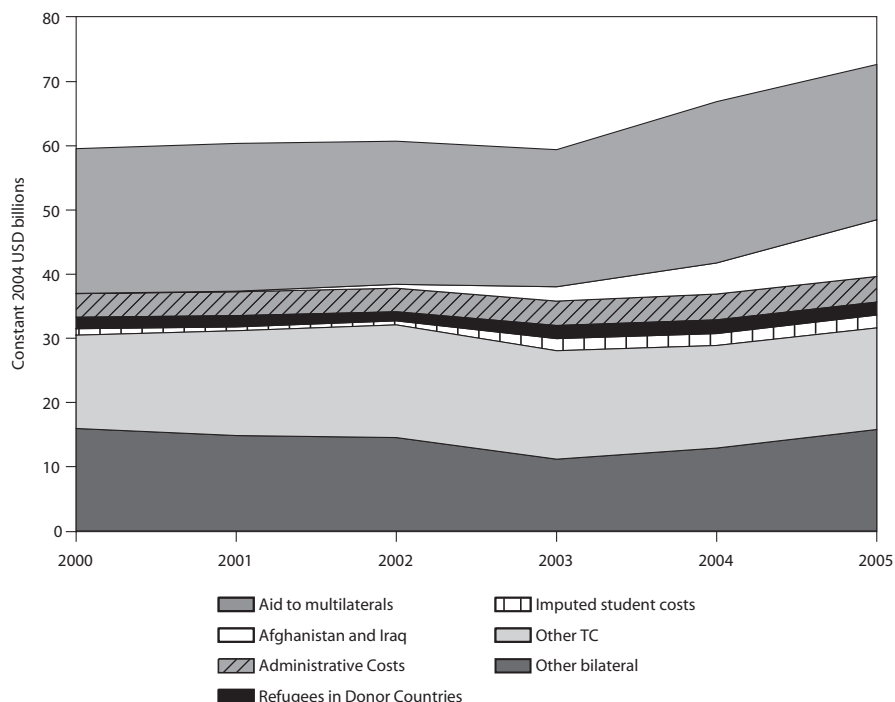
Figure 1
Aid components, 2000 2005. Source: OECD Development Aid Committee (DAC) data.

harmonization and coordination are vital to reduce the currently high transaction costs of aid, which divert scarce resources   including the time demands on the most competent civil servants   into projects and activities that have too often been donor priorities rather than recipient national priorities.

Considerable progress has been made in harmonizing approaches to aid. The 2005 Paris conference and subsequent Declaration on Aid Effectiveness brings together multilateral and bilateral donor agencies around common standards. While the MDGs provided a landmark in terms of defining common goalposts for development, the harmonization agenda aimed to ensure that the donors played as one team. Significant improvements in coordination among a number of the traditional major donors point the way forward. As diverse aid flows are increasingly combined into broad multidonor activities, the challenge for donors is

to convince their skeptical voters that their taxes have been spent wisely. Although attributing the impact of aid to individual donors may be counterproductive, in that it suggests that development is not a national responsibility, aid   like all public expenditures   requires accountability.

Levels of Aid, Aid Quality, and Evaluation Aid flows increased markedly between 1945 and 1960, but subsequently slowed and from 1990, when it averaged 0.34 percent of the gross national income (GNI) of the donor countries, declined to around 0.22 percent of the GNI of high-income countries in 2001, as shown in figure 2. The recommitment to aid at the Monterrey Conference in 2002 finally arrested this decline, with ODA reaching a record high in 2005. Subsequently, aid flows again slipped back, reflecting the fact that the inclusion of Iraq and Nigerian debt write-offs had temporarily inflated the numbers earlier. Going forward, the key uncer-
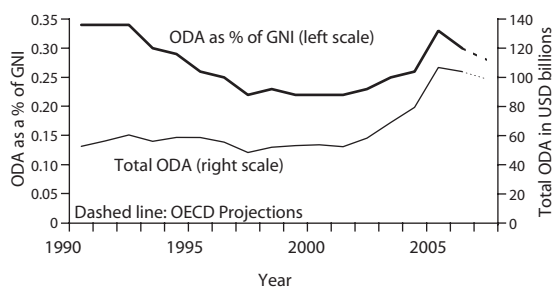
Figure 2

Aid flows. Source: OECD Development Aid Committee (DAC) data.

Figure 3

Foreign aid receipts as a percentage of low and middle income GNI, 1960 2006. Source: OECD Development Aid Committee (DAC) data.

tainty is the extent to which donors will honor the pledges made at Monterrey and reiterated at the 2005 Gleneagles G8 Summit, where it was agreed to double aid to Africa and to achieve the 0.7 percent commitment.

In comparison with domestic investment and government expenditure, aid flows are small. Development aid to developing countries in 2005 was a record $107 billion, but even without taking account of the fact that most of this was not transferred to developing countries, aid remained less than one-third of foreign direct investment in developing countries ($334 billion), which itself was only a small fraction of total investment in developing countries (more than $2 trillion). Although increasing the volume of aid is vital, improving its quality is even more important.

Both the relatively small levels of aid and the instability of these flows for middle- and low-income countries are reflected in figure 3. Since aid is typically small compared to budget and private investment flows, and more unstable, the key challenge is to ensure that aid effects systemic changes such as introducing ideas and improving practices. These in turn have positive effects on growth and poverty reduction. Such indirect effects are difficult to measure and attribute, however, and, as indicated earlier, attribution risks undermining government leadership and harmonization. It is important that aid be provided not solely for short-term and relatively easily measurable objectives.

A growing body of literature had sought to use randomized techniques to provide improved rigor to the evaluation of aid. Impact evaluation and other evaluative methods may be expected to become a necessary part of the toolbox of aid agencies as they and the recipient governments seek to enhance the effectiveness of pubic expenditures. These techniques yield powerful insights when applied through carefully designed studies of projects or programs but typically they are less helpful at the macro policy level, where there is limited scope for randomization or associating single actions or actors with particular outcomes.

The complexity of social and economic change means that the impact of foreign aid cannot be easily separated from other factors. The most successful projects are ones in which recipients are strongly committed and in charge of the development process, and where there is good partnership among donors and with local leadership. Recipients are then particularly well placed to draw on foreign aid, with a view to both learning lessons from international experience and funding well-defined investments that can underpin growth and poverty reduction. For such recipients, the objective is to become self-sustaining and to turn from being net aid recipients to donors, as has been the experience of China.

The increasing focus of aid on countries with relatively sound policies has meant that the estimated

poverty-reduction productivity of ODA is significantly better than it was in the early 1990s. Countries with poor policies and poor institutional quality cannot simply be isolated, however. Some recent studies have found that the impact of aid on growth is strong regardless of institutions and policies, and it is necessary to be cautious about the evidence in this area.

Weak States If viable institutions and policies are conducive to effective use of aid, what should be done in weak or failed states where these do not exist? Such states usually lack the governance, institutions, and necessary leadership for successful reform. Each incompetent government has its own specific problems, and aid interventions must be tailored to addressing and overcoming these. A growing specialist literature has examined the lessons of interventions in weak states, and a number of pointers to supporting such countries have emerged. It has been shown that knowledge transfer and capacity building are more effective than large-scale financial transfers and that improving basic services such as health and education is also important. In such situations, the potential role for the United Nations, in particular in terms of coordination, is often underestimated and tends to be inadequately resourced in terms of personnel and funds. While Ethiopia, Mozambique, and Uganda are postconflict success stories, unfortunately many other countries have seen little significant progress in the past decade. The challenge for aid agencies is to remain engaged but not to throw good money after bad policies. Part of the response lies in attention to coherence between aid and other policies, especially in weak states and in countries beset by conflict, as emphasized by Collier (2007).

Global Public Goods Although much of the attention in the aid debate has been on support to countries and national projects, global initiatives are also significant beneficiaries of aid. An important step in development is supporting what are known as global public goods. These are public goods whose benefits are felt beyond the border of any one country, benefiting the poor in many countries and even all humanity. The pooling of resources and coordination across national boundaries, and between the public and private sectors, is vital in addressing such global issues. These include combating major infectious diseases such as HIV/AIDS and malaria, developing better crops, managing intellectual property, and dealing with climate change.

An example of a global aid initiative is the Consultative Group on International Agricultural Research (CGIAR). This is a partnership working toward sustainable food security and poverty reduction through scientific research in the fields of agriculture, forestry, fisheries, policy, and environment. The partnership includes countries, international and regional organizations, private foundations, national agricultural research systems, civil society, and the private sector, all of which support the work of 15 international research centers. The CGIAR significantly helped the Green Revolution, which began in South Asia in the 1970s and has led to impressive gains in production of basic food crops across the developing world. Between 1970 and 1997, yields of cereals in developing countries rose more than 75 percent, coarse grains 73 percent, root crops 24 percent, and pulses nearly 11 percent.

Private Initiatives and New Donors The Global Alliance for Vaccines is an example of aid in support of a global public good. It is also indicative of the rapidly growing contribution of private donors. The Bill and Melinda Gates Foundation and other relatively recently established private aid donors, as well as privately launched campaigns such as the Clinton Global Initiative, are adding to the contribution and in terms of financial flows overtaking the more traditional nongovernmental aid institutions such as the Ford and Rockefeller Foundations, civil society groups such as Oxfam and Save the Children, and religious foundations. Together, these nongovernmental flows are channeling aid that in many countries, and perhaps also in aggregate, now exceeds the public flows. This greatly increases the potential as well as the complexity of aid (see Klein and Harford 2005). In addition to the rapid growth of private flows, interesting experiments are under way that seek to build on public-private partnerships and leverage official flows. One such example is the pilot Innovative Funding Facility for Immunization. The

French-led actions to increase aid flows through a levy on airline travel and the British-led securitization of future aid flows to raise finance from capital markets are indicative of the way that innovation and partnership are beginning to change the shape and potential of the traditional aid architecture.

The transformation of traditional aid recipients, notably China, into highly significant aid donors has challenged the aid establishment. The thrust of the OECD official donors has been closer coordination and ensuring that the recipients conform to governance and other standards. China's and other new aid donors' perspectives and objectives are not necessarily aligned with those of the traditional donors, and these new donors have not been part of the OECD Development Aid Committee coordination process. The growth in these new aid flows, both from private foundations and from governments such as China and Venezuela, has meant that there have been significant increases in the volume of aid. The key questions concern the coordination and quality of this aid and the extent to which it will be able to avoid the dangers of tied aid and aid fragmentation that undermined previous surges in aid. The ending of the Cold War and the ability of countries to access aid together with a range of development ideas and technologies have made the biggest difference to development outcomes in recent years.

A New Way Forward The polarization of the aid debate between the strong advocates, such as Sachs (2005), and deep skeptics, such as Easterly (2006), reflects an oversimplification of the complexity of the development challenges and the need for strong domestic and international actions. Resources alone will not be sufficient to ensure that poverty goals are met. Level of commitment and quality of policies and institutions in recipient countries are the primary determinants of progress. It is also evident that, when a country is committed to reform and poverty reduction, external support (which may not be limited to aid) can have substantial payoffs. An important area in which rich countries can provide support is through reforms of their own trade and other policies. As important as robust global growth is reform of the protectionist policies of rich countries (such as

in agriculture and textiles), which are so damaging to poorer countries. As Goldin and Reinert have shown (2007), changes in trade, investment, migration, environment, security, and technology policies in rich countries would pull many out of poverty. The Commitment to Development Index and associated analysis by the Center for Global Development provides an indicative summary of the importance of coherence between aid and other policies.

Coherence between aid policies and other policies is vital. For example, aid donors' support for health and educational systems is undermined by the recruitment of teachers, doctors, and nurses to work in the rich countries. Similarly, support for agricultural programs is undermined by protectionism and subsidies that prevent developing countries from competing on world markets. With continued reform momentum and steady external support, past experience suggests that developing countries can extend and deepen the progress of the last half-century. Despite the progress made in the past 50 years, an immense poverty challenge remains. Approximately 1 billion people still live on less than one dollar per day. Aid has never been more effective in supporting growth and reducing poverty, and returns on aid have increased sharply. Raising the volume and the quality of aid is a moral, strategic, and economic imperative.

See also aid, bilateral; aid, food; aid, humanitarian; aid, military; evolution of development thinking; Millennium Development Goals

FURTHER READING

Collier, P. 2007. *The Bottom Billion: Why the Poorest Countries Are Failing and What Can Be Done About It.* Oxford: Oxford University Press. Seeks to identify the primary processes trapping the world's poorest people and suggests means to overcome their predicament.

Collier, P., and D. Dollar. 2004. "Development Effective ness: What Have We Learnt?" *The Economic Journal* 114 (496): F244 F271(1). An overview and assessment of aid that seeks to identify determinants of aid effectiveness.

Easterly, W. 2006. *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little*

Good. New York: Penguin. Condemns top down aid planning as a failure of utopian social engineering.

Goldin, I., and O. Knudsen. 1993. *Trade Liberalization: Global Economic Implications.* Paris and Washington, DC: OECD and World Bank. Explores the implications of trade protectionism and potential for trade reform to enhance development.

Goldin, I. and K. Reinert. 2007. *Globalization for Devel opment: Trade, Finance, Aid, Migration and Policy.* Washington, DC: World Bank and Palgrave Macmillan. Identifies key economic flows and recommends policies to ensure that globalization supports development.

Goldin, I., H. Rogers, and N. Stern. 2002. *A Case for Aid: Building a Consensus for Development Assistance.* Wash ington, DC: World Bank. Evaluates evidence on aid and includes readings related to the MDGs.

Hansen, H., and F. Tarp. 2000. "Aid Effectiveness Dis puted." *Journal of International Development* 12 (3): 375 98. Examines the literature on the aid growth re lationship, concluding aid works even in countries hampered by unfavorable policy environments.

Kaul, I., I. Grunberg, and M. A. Stern. 1999. *Global Public Goods: International Cooperation in the 21st Century.* Oxford: Oxford University Press. UNDP edited col lection of papers arguing that development cooperation has moved beyond aid and poverty reduction toward wider provision of global public goods.

Klein, M., and T. Harford. 2005. *The Market for Aid.* Washington DC: IFC. Introductory account of the me chanics and rapid evolution of the global aid industry.

OECD, Development Assistance Committee (DAC). Various Years. *Development Co operation Report.* Paris: Organisation for Economic Co operation and Devel opment. The authoritative source of information on the foreign aid policies and programs of donor countries, published annually.

Reinert, K. 2005. *Windows on the World Economy: An In troduction to International Economics.* Mason, Ohio: South Western Thomson. Provides an analytic under standing of the concepts of international trade, pro duction, finance, and development.

Riddell, R. 2007. *Does Foreign Aid Really Work?* Oxford: Oxford University Press. An extensive overview of aid, examines where aid has failed and why, and identifies practical changes needed to make aid effective.

Sachs, J. 2005. *The End of Poverty: Economic Possibilities for Our Time.* New York: Penguin. Argues extreme poverty can be ended by 2025 through nine specific steps to help the poorest onto the first rung of the "development ladder."

Sen, A. 1982. *Poverty and Famines: An Essay on Entitle ment and Deprivation.* Oxford: Clarendon Press. De velops an "entitlement approach" to development, fo cusing on the causation of starvation, and a general analysis of the characterization and measurement of poverty.

———. 1999. *Development as Freedom.* New York: Knopf. Argues that open dialogue, civil freedoms, and politi cal liberties are vital dimensions of sustainable develop ment.

Szirmai, A. 2005. *The Dynamics of Socio Economic Devel opment: An Introduction.* Cambridge: Cambridge Uni versity Press. A nontechnical introduction to develop ment studies, exploring the dynamics of development and stagnation in developing countries.

Taylor, L. 1993. *The Rocky Road to Reform: Adjustment, Income Distribution, and Growth in the Developing World.* Cambridge, MA: MIT Press. A collection of case studies set in a broader context of examining why nations grow at different rates with inequitable patterns of wealth and income distribution.

United Nations Conference on Trade and Development (UNCTAD). 2006. *World Investment Report.* Geneva: United Nations. Provides global trends in foreign direct investment and identifies measures to improve invest ments contribution to development.

World Bank. 1994. *Reducing the Debt Burden of Poor Countries: A Framework for Action.* Development in Practice Series. Washington, DC: World Bank. Early discussion of the policies which aimed to reduce debt subject to performance improvements in highly in debted countries.

———. 2005. *World Development Report 2006: Equity and Development.* Washington, DC: World Bank. Con cludes that inequality of opportunity sustains extreme deprivation and recommends ensuring more equitable access to health care, jobs, capital, secure land rights, and political freedoms.

IAN GOLDIN

■ aid, international, and political economy

International donors provided more than U.S. $2.4 trillion in the form of foreign aid from 1960 to 2004. In many African countries, for example, the size of the official development assistance (ODA) divided by gross domestic product (GDP) exceeded 10 percent during that period. Despite this large amount of financial aid, the economic performance of the recipient countries has been disappointing. The poor macroeconomic impact of foreign aid raises questions about its efficiency. Political economy considerations can provide explanations for the scarce effectiveness of foreign aid. By the term *political economy* we mean the allocation of markets and other economic institutions subject to the political environment in which these institutions work. In particular, the process of decision and delivery of foreign aid present multiple agency problems, derived from the existence of asymmetric information (the donor and recipient do not have access to the same information) and moral hazard (receipt of aid causes the recipient to exert less effort to solve its own problems), and inappropriate institutional designs.

Most of the academic research finds that international aid is ineffective in fostering economic growth. For instance, Boone (1996) finds that aid does not significantly increase investment or any human development indicator, but it increases the size of the government. Burnside and Dollar (2000) find that aid works, but only when there are good fiscal, monetary, and trade policies. Easterly, Levine, and Roodman (2003) dispute these results, however, since they are sensitive to changes in the sample period or filling in the missing observations.

One of the basic reasons that development aid does not seem to work is the presence of perverse incentives and faulty institutional designs. The political economy perspective treats foreign aid as the result of the incentives and motivations of the donors, the response of country recipients, and the interaction between donors and recipients.

The Incentives of Donors and Aid Agencies

Efficient public programs are supported by the existence of sufficient (perfect) information and the accountability of the representatives. Households in developed countries are, at the same time, clients and citizens. As clients, they have information about the programs intended for their benefit and as citizens in democratic countries they can hold politicians accountable for their poor performance. In the case of foreign aid this feedback loop is absent: taxpayers in donor countries (citizens) are not the beneficiaries of the aid, and they do not have information about its effectiveness; and the intended beneficiaries (clients) do not have the right to vote in donors' countries and, therefore, cannot affect the politicians in charge of approving the aid programs. We can describe this situation as a broken information feedback loop (Martens et al. 2002). In the absence of effective feedback from the intended beneficiaries of aid, the interests of consultants and suppliers of goods shape many decisions of aid agencies: these groups are the direct beneficiaries (control the information about the programs) and have direct influence on the decision-makers of donor countries.

Donors and aid agencies also have multiple objectives and represent multiple principals that often do not share the same objectives. For these reasons the institutional design of official aid agencies is difficult even if donors have good intentions. First, the multiplicity of objectives and the difficulty of performance measurement make it complicated to link the salaries of workers in aid agencies to their performance. Therefore, aid bureaucracies tend to define their output in terms of money disbursed (or volume of loans) instead of the rate of poverty reduction or economic development of recipient countries (Easterly 2003). Second, aid bureaucracies tend to develop their own procedures and increase the complexity of tenders and contracts (see Easterly 2006). Finally, aid agencies will tend to choose projects that minimize the risk of bad publicity, not the ones that maximize, for instance, the reduction of poverty.

Donor countries may also have objectives that undermine the effectiveness of foreign aid. In fact, almost half of the foreign aid provided by the members of the Organisation for Economic Co-operation and Development does not have as basic objectives poverty alleviation and economic

development. Alesina and Dollar (2000) find that while some donors have the right incentives (raising income level, poverty reduction, institutional improvement), many others are driven by political and strategic considerations. The wrong incentives include targeting aid to inefficient and nondemocratic former colonies or developing countries that vote in favor of the proposals of the donor at the UN. The strategic interest in the Middle East explains why Egypt and Israel receive so much economic support from Western countries.

Another possible reason for the ineffectiveness of foreign aid is the so-called warm glow effect. Donor citizens may be interested only in the fact of giving itself and not in its effects on the recipient countries. Therefore, the utility of donors does not depend on the well-being of the citizen of recipient countries but rather on the satisfaction the donors derive from providing aid to developing countries. This interpretation is consistent with the evaluation of policies based on inputs (money disbursed) rather than on outcomes (effect of the aid on poverty and economic development). It is also consistent with voters who are worried about the total amount of money that is dedicated to foreign aid (say 0.7 percent of GDP) and not about the results that aid produces in recipient countries.

Finally, since there are multiple principals (donors and agencies), the effectiveness of foreign aid depends also on the degree of coordination among them. The need for coordination has increased over time because the level of fractionalization, or diversity, of donors (many donors with a small share of aid) has increased from 0.25 in 1960 to 0.68 in 2003. Fractionalization is measured as 1 minus the sum of the squares of the shares of different donors. Empirical evidence suggests that higher donor fragmentation is associated with a reduction in the bureaucratic quality in recipient countries.

The Interaction between Donors and Country Recipients The relationship between donors and country recipients is usually subject to principal-agent problems and moral hazard. The relationship between a principal (the donor) and an agent (the country recipient) can lead to problems when the interests of the principal and the agent do not coincide and the principal has less than perfect information about the actions of the agent. The usual game theoretical solution to these types of problems is to write a contract, or design an incentive mechanism, to force the agent to perform as the principal wants. This incentive mechanism usually takes the form of aid conditionality: the requests associated with the concession of some types of foreign aid. The credibility of the threat to retire aid if the conditions of development of the projects are not satisfied is critical for the effectiveness of the conditionality, however. Recipient countries realize that these threats are not credible because donors do not have strong commitment ability. Donors may want to punish countries that do not fulfill their promises but, at the same time, they are pressed by their own constituencies to continue helping poor countries. The problem is reinforced by the fact that the status of the managers within the aid agencies is determined by the size of their budget and the disbursement of funds. Empirical studies show this lack of commitment technology. For instance, Burnside and Dollar (2000) do not find evidence of a significant positive relationship between aid flows (ODA) and good policies. Other studies find that there is no relationship between the developing countries' score on quality of governance (corruption, democracy, service delivery, etc.) and who receives adjustment loans, for instance, from the World Bank.

Some authors have proposed alternative institutional designs to overcome the lack of credibility of donor threats. Svensson (2000) shows that, when there is no commitment technology available for donors, the delegation of the aid budget to an institution with less aversion to poverty (underdevelopment) than the donor will improve the welfare of the citizens of the recipient country since it makes the threat credible. This theoretical result has yet to be implemented, however. If the aid game implies a donor and several recipients' countries, or a pool of projects that are disbursed to each individual country depending on its relative performance, it is possible to design an aid tournament in which potential aid recipients compete for funds. In a bilateral relation-

ship the recipient country has more bargaining power and, therefore, the credibility of the donor's threats is weak. If the donor can choose between different recipients, however, then recipient countries have to commit to a high level of reform effort if they want to win the aid tournament. In fact, donors can impose a cost on the recipients (for instance, they may ask for a set of reforms even to get considered in the tournament). This is the traditional "stick and carrot" type of strategy. In addition, having alternative countries to receive the aid of a donor reduces the ability of one particular recipient country to use hold-up strategies.

Another problem associated with the moral hazard present in the relationship between donors and developing countries is the so-called Samaritan's dilemma (Gibson et al. 2005), which can be represented as a two-person game. The donor (Samaritan) wants to help a poor country to get out of poverty. Therefore, it gets a higher payoff from helping, or giving a large amount of help, than from not helping (or providing less aid). The recipient has to decide how much effort to exert. Obviously, the donor prefers the big-effort action. The recipient country, however, prefers to exert little effort. If the game is repeated, as happens in real interactions between donors and developing countries, then we get little effort and high levels of foreign aid. There are many examples of this situation. Donors give large amounts of food aid to developing countries with food crises. Recipient countries know this and, therefore, they have incentives to relocate their own resources away from agriculture. This action will increase the likelihood of food shortages in the recipient country, which will generate constant food aid from the donor. Then the recipient country produces less food than necessary, and the developed country sends food aid on a regular basis.

The solution to this principal-agent problem is to monitor the recipient country. However, the lack of information and the absence of accountability of the bureaucracies of recipient countries make monitoring a nonviable solution. Another possible solution for this type of problem is to threaten to withdraw the aid in the future. If this threat is credible

it can support the best outcome (high level of aid and high level of effort, which economists describe as a Pareto efficient equilibrium). In game theoretical terms the donor threatens the recipient country with ending in the worst situation for both agents (low level of aid and low level of effort, which economists call, in this case, Nash equilibrium). However, donors are usually unable to make this type of credible threat.

The Delivery of Aid in the Recipient Countries
The distribution of aid at the operational level is also subject to a principal-agent relationship and to a moral hazard problem. The moral hazard problem exists because the benefit of the contract to the donor depends on the actions taken by the recipient, and the recipient may not have incentives to behave properly after the donor has provided the funds. The recipient will take the action most beneficial to his own preferences. Therefore, the terms of the contract must be modified to provide the recipient enough motivation to behave properly. Bureaucracies are even more important in recipients' countries than in donors' agencies. The civil servants of the public bureaucracies of developing countries have a long-term contract with a low wage. Public regulation makes it difficult to fire government workers even if their performance is not appropriate. One particularly important consequence of this lack of incentives in the provision of education and health in developing countries is the so-called missing doctors and teachers. For example, Banerjee and Duflo (2006) report that, in India, the absence rate for teachers is more than 24 percent and for health providers is more than 40 percent. This is the usual shirking behavior predicted by the principal-agent theory.

In addition, since salaries for workers in the public bureaucracies of developing countries are low, they tend to run private moneymaking activities or extract side payments from citizens (extort payment for free health services, demand money for a license, etc.). Corruption and rent-seeking behavior are often the result of severe moral hazard problems combined with the absence of clearly defined property rights. One important explanation for the lack of effect of foreign aid on growth and social development is the

capture of aid by government officials and politicians. The measurement of the extent of corruption in the allocation of foreign aid in developing countries is difficult because accounting systems work poorly and local accountability is weak. The World Bank has used Public Expenditure Tracing Surveys (PETS) to calculate how much money is diverted from the time the central government receives the funds until they reach the final beneficiaries. The proportion of funds that reach the intended beneficiaries tends to be very low. For instance, Reinikka and Svensson (2004) analyze a large public educational program (a capitation grant to cover schools' nonwage expenditures) in Uganda financed by the central government using district offices for its distribution. They find that many schools received nothing. On average, schools got only 13 percent of the government spending in the program. In Tanzania the analysis of the Primary Education Development Plan launched in 2002 by the World Bank and bilateral donors shows that, on average, only 20 percent of the funds disbursed at the central level finally reached the schools.

Rent-seeking activities may even produce negative consequences in the recipient countries. Djankov, Montalvo, and Reynal-Querol (2006) provide empirical evidence that a sudden windfall of resources in the form of foreign aid can damage the political institutions of the receiving country by reducing checks and balances in government and democratic rules. They label this effect the curse of aid. Other authors provide evidence that Somalia's civil war was caused by the desire of different factions to control the large amount of food aid that the country received.

The Market for Aid Some authors have argued that a possible solution to the agency problems in the aid industry is to increase competition among aid agencies (Klein and Harford 2005) and recipient countries (Svensson 2003) and use competition to make the market work better. Aid agencies should compete to fund the best projects, enabling donors to channel funds through the agencies that produce the best results; and developing countries should compete to receive official aid. This increase in competition, if accompanied with improved monitoring and benchmarking of aid agencies and proper evaluation of outcomes for the projects, may alleviate some of the incentive problems originated by the multiple layers of delegation in the delivery of foreign aid.

Notice that this does not contradict the need for coordination among the donors. Competition is good in the sense that recipient countries will choose the most efficient project, and therefore will end up with the best donor for them. So, the existence of many donors competing for recipients does not imply that many donors will be delivering money to the same project, but only the most efficient one.

See also aid, international

FURTHER READING

Alesina, Alberto, and David Dollar. 2000. "Who Gives Foreign Aid to Whom and Why." *Journal of Economic Growth* 5: 33 63. Studies the pattern of allocation of foreign aid from various donors to receiving countries.

Banerjee, Abhijit, and Esther Duflo. 2006. "Addressing Absence." *Journal of Economic Perspectives* 20 (1): 117 32. Presents evidence on a number of innovative strategies to reduce absenteeism in government and nongovernmental organization run schools and health facilities.

Boone, Peter. 1996. "Politics and the Effectiveness of Foreign Aid." *European Economic Review* 40: 289 329. Finds that aid does not significantly increase investment and growth, nor benefit the poor as measured by improvements in human development indicators, but does increase the size of government.

Burnside, C., and D. Dollar. 2000. "Aid, Policies, and Growth." *American Economic Review* 90 (4) (September): 847 88. One of the most influential papers on the recent literature of the effectiveness of foreign aid. It finds that aid has a positive impact on growth in developing countries with good fiscal, monetary, and trade policies but has little effect in the presence of poor policies.

Collier, Paul, and David Dollar. 2002. "Aid Allocation and Poverty Reduction." *European Economic Review* 46: 1475 1500. Reallocating aid is politically difficult, but it may be considerably less difficult than quadrupling aid budgets, which is what the authors estimate would be necessary to achieve the same impact on poverty reduction with existing aid allocations.

Djankov, Simeon, José G. Montalvo, and Marta Reynal Querol. 2006. "Does Foreign Aid Help?" *Cato Journal* 26 (1): 1 28. Concludes that improving the effectiveness of aid requires increasing the responsibility of aid re ceiving countries (by providing loans instead of grants in a credible environment), reducing the cost of remit tances to developing countries, and improving the co ordination among donors.

Easterly, William. 2003. "The Cartel of Good Intentions: The Problem of Bureaucracy in Foreign Aid." *Journal of Policy Reform* 5 (2): 67 91. Concludes that policymakers in rich and poor countries should experiment with de centralized markets to match those who want to help the poor with the poor themselves freely expressing their needs and aspirations.

———. 2006. *The White Man's Burden.* New York: Pen guin Press. Should be compulsory reading for anybody interested in economic development.

Easterly, William, Ross Levine, and David Roodman. 2003. "Aid, Policies, and Growth: A Comment." *American Economic Review* 94 (3): 774 80. Finds that the results of Burnside and Dollar (2000) are not robust to extending the sample period, nor to filling in missing data for the original period 1970 93.

Gibson, Clark, Krister Andersson, Elinor Ostrom, and Sujai Shivakumar. 2005. *The Samaritan's Dilemma: The Po litical Economy of Development Aid.* Oxford: Oxford University Press. Suggests ways to improve aid's effec tiveness.

Klein, Michael, and Tim Harford. 2005. *The Market for Aid.* Washington: International Financial Corporation. Argues for rigorous methods of evaluation and creative use of the private sector to produce a more effective aid industry that encourages new experiments.

Martens, Bertin, Uwe Mummert, Peter Murrell, and Paul Seabright. 2002. *The Institutional Economics of Foreign Aid.* Cambridge: Cambridge University Press. Examines the institutions, incentives, and constraints that guide the behavior of people and organizations involved in the implementation of foreign aid programs.

Reinikka, Ritva, and Jakob Svensson. 2004. "Local Capture: Evidence from a Central Government Transfer Program in Uganda." *Quarterly Journal of Economics* 119 (2): 679 705. Using panel data from a unique survey of primary schools, the paper assesses the extent to which government grants actually reach the intended end user (in this case schools). The survey data reveal that during 1991 95, the schools, on average, received only 13 percent of the grants.

Svensson, Jakob. 2000. "When Is Foreign Aid Credible? Aid Dependence and Conditionality." *Journal of Development Economics* 61: 61 84. Shows that an aid contract, as proposed in the literature, is not time consistent. The paper also provides some evidence supporting the basic idea that aid induces weak fiscal discipline and that increased fiscal difficulties lead to higher inflow of aid.

———. 2003. "Why Conditional Aid Does Not Work and What Can Be Done About It?" *Journal of Development Economics* 70: 381 402. Studies a simple reform that introduces ex post incentives for the donor to reward good policies.

JOSÉ G. MONTALVO

■ aid, military

Military aid can take many different forms, includ ing grants, loans, or credits to purchase defense equipment, services, and training. Often military aid is "tied," in that recipients must use the funds to buy defense goods and services from the donor. The aim of military aid is to assist recipients with a variety of security problems. The recipient countries may face a number of different security threats, such as inter national war, internal rebellion, or terrorism. Military assistance is often supplied to help not only with the recipient's national security but with regional and global security threats. These international consider ations make the study of military aid controversial.

Military aid may imply that this is part of a do nor's official development assistance (ODA). How ever, *aid* as defined by the Organisation for Eco nomic Co-operation and Development (OECD) explicitly excludes military aid. The OECD defini tion of ODA to developing countries includes grants or loans to countries that are undertaken by the of ficial sector in order to promote economic develop ment and welfare. Grants, loans, and credits for military purposes are excluded.

Rather than being part of a donor's aid budget, military aid is part of the defense budget. Defense expenditure is typically one of the least reliable components of a government's budget. The level and composition of military expenditures are often treated as state secrets and sizable portions are not open to public scrutiny. Generally, it is hard to study the effect of military aid since data are scarce and unreliable (Brzoska 1995).

Military and Development Aid Military and development aid are both targeted at foreign nations. An interesting question is how much of foreign assistance is allocated to the seemingly different purposes of overseas security on the one hand and development on the other. This analysis requires data on the different components of aid. Development aid data are available through national governments and the OECD. Few countries, however, make military aid data available, the United States being a notable exception. When one sums up U.S. military and development aid to "total aid," it is interesting to see how the composition of total aid changed over time. From 1970 until 1975 the share of military aid was larger than the share of development aid; it accounted for 53 percent of the total aid figure. After the end of the Vietnam War the composition of total aid changed, and by the early 2000s the share of military aid in total aid was only 23 percent.

Military aid data for the other major donors are not readily available but if the shares of military and development aid are similar to those of the United States, a rough estimate of global military aid can be calculated. In 2006 the sum of all development aid was U.S. $85 billion; therefore, globally military aid was probably around $26 billion. Not all of this military aid is provided to poor countries. Much of this military aid goes to developed countries such as Israel and Turkey. A large proportion of low-income countries' military expenditure is provided through foreign assistance, however. Since this is precisely the aim of military aid, there does not seem to be any problem with financing other countries' military budgets. But is all of the development aid used for the purpose of economic development for which it is intended or is some of it leaking into military spending, which is explicitly excluded from the definition of development aid?

A large number of studies examine what determines the allocation of aid to the different recipient countries (Alesina and Dollar 2000). The literature distinguishes between recipient need, recipient merit, and donor self-interest as motivations for development aid. It seems that, for example, U.S. aid allocation over the past few decades has been predominantly driven by geostrategic interests rather than by recipient needs (McKinley and Little 1979). In this sense donors use development aid for purposes other than assisting economic development in the recipient country. If donors do not always use development aid for its stated purposes, what about the recipient countries? In the development aid literature there is a large debate on fungibility   that is, there is concern that funds that are earmarked for a specific purpose are being used to finance projects and programs for which they were not originally intended (Feyzioglu, Swaroop, and Zhu 1998). It has been estimated that about 11 percent of development aid leaks into military expenditure (Collier and Hoeffler 2007). In 2006 this would have amounted to about $9 billion. Taking this leakage into consideration, we can state that the share of military aid in total aid is greater than the officially reported 23 percent; it is about 32 percent.

Military Spending in Low-Income Countries Is this explicit and implicit funding of military expenditure in low-income countries a problem? After all, these countries have security concerns, and without security no state can provide basic services to its citizens. Without security, economic development is not possible. Studies show that higher military expenditure in poor countries hinders their development (Deger and Smith 1983), suggesting that these countries should spend more on basic services such as infrastructure, health, and education rather than on the military. There is also evidence that higher military expenditure does not lower the risk of civil war, which is the most common form of large-scale violent conflict. It may even be the case that higher military expenditure in postconflict situations is particularly damaging because it increases the risk

of a recurrence of civil war. Rebels may not trust the government to stick to a peace settlement if they observe an increase in armament (Collier and Hoeffler 2006). They are then more likely to restart a rebellion. To summarize, many developing countries receive direct and indirect military aid, but the evidence so far suggests that it neither helps them to develop economically nor increases their security.

See also aid, bilateral; aid, international; aid, international, and political economy

FURTHER READING

Alesina, A., and D. Dollar. 2000. "Who Gives Foreign Aid to Whom and Why?" *Journal of Economic Growth* 5: 33 63. Empirical examination of how donors allocate aid to poor countries.

Brzoska, M. 1995. "World Military Expenditures." In *Handbook of Defense Economics*, vol. 1, edited by K. Hartley and T. Sandler. Elsevier: North Holland, 45 67. Overview of the definition of military expenditure and the difficulties in compiling statistics across countries and time.

Collier, P., and A. Hoeffler. 2007. "Unintended Consequences: Does Aid Promote Arms Races?" *Oxford Bulletin of Economics and Statistics* 69: 1 28. Provides a general estimation of the demand for military expenditure and how aid changes this relationship.

———. 2006. "Military Expenditure in Post Conflict Societies." *Economics of Governance* 7: 89 107. Theoretical and empirical analysis of the impact of military expenditure in postconflict situations.

Deger, S., and R. Smith. 1983. "Military Expenditure and Growth in Less Developed Countries." *Journal of Conflict Resolution* 27: 335 53. Empirical cross country analysis of how military expenditure affects growth.

Feyzioglu, T., V. Swaroop, and M. Zhu. 1998. "A Panel Data Analysis of the Fungibility of Foreign Aid." *World Bank Economic Review* 12: 29 58. Empirical examination of the use of foreign aid.

McKinley, R. D., and R. Little. 1979. "The US Aid Relationship: A Test of the Recipient Need and the Donor Interest Models." *Political Studies* 27: 236 50. Statistical analysis of how the United States distributes foreign aid.

Organisation for Economic Co operation and Development. http://www.oecd.org/dataoecd/42/30/1860571 .gif. A source for data about overseas development aid.

Stockholm International Peace Research Institute. http:// www.sipri.org/contents/milap/. A source for data on worldwide military expenditures.

U.S. Census Bureau. http://www.census.gov/compendia/ statab/foreign commerce aid/foreign aid/. A good source for data about American military aid.

PAUL COLLIER AND ANKE HOEFFLER

■ **aid for trade**
See trade-related capacity building

■ **air transportation**

Air transportation is part of a service supply chain that moves both passengers and cargo. Although commercial airlines offer the immediate form of transportation, they are tied to airports for their takeoffs and landings and to air navigation services for guidance and control in route. Passengers and cargo also need to move to and from the airport. Air transportation, therefore, is a sector that caters to the transportation of passengers overall but especially those going the shortest distances and to the movement of high-value, low-bulk freight.

Air transportation is a major industry in its own right and acts as a lubricant for many other economic activities. In broad terms, air transportation accounts for about 1 percent of the gross domestic product of the European Union and slightly less for the United States economy. Globally, in 2005 there were about 18,000 commercial aircraft, carrying 1.6 billion passengers and more than 43 million tons of freight annually, and serving more than 10,000 airports. These statistics, however, ignore the role that air transportation can play in stimulating economic growth and trade (it is estimated that between 35 and 40 percent of international trade by value is moved by air transportation), linking together diverse communities, and fostering particular types of industry

such as tourism and specialized agricultural sectors such as exotic fruits and flowers.

Although air transportation played a role in providing mail service, and some limited passenger and cargo service before 1939, it has gained significance only since the end of World War II with major technological advances in aircraft design and enhanced air traffic navigation systems. The advent of commercial jet aircraft in the 1950s provided faster and cheaper long-haul services, and in the 1960s the development of wide-bodied aircraft increased carrying capacity, which cut the costs of air travel considerably, reduced environmental concerns, and improved safety. Radar and improved communications, together with institutional changes and accompanying new management practices beginning in the late 1970s, resulted in further cost reductions and the expansion of airline service networks.

Modern scheduled air transportation networks take a variety of forms. To meet social or political objectives, governments stipulate certain types of service that must be supplied either by state-owned airlines or through financial support to private airlines from the taxpayer. Although government interventions in the market are considerable, they have diminished in many countries since the 1980s. As a consequence, profit is the motivating factor for an increasing portion of the air transportation system worldwide.

Where the market plays a dominant role, three broad types of scheduled service are common. Many services are essentially point-to-point with the airline serving a set of individual origins and destinations, often in a linear network akin to a bus service. This approach may be refined into radial networks with a carrier concentrating services from a base airport but offering no coordinated services involving a change of plane that allow passengers who want to continue on to another destination from that base. Finally, hub-and-spoke operations involve a major airport serving as an interchange facility (a little like a postal sorting office) that consolidates passengers or freight from diverse origins on flights to a range of destinations. This type of operation allows economies of scope (the ability to offer a range of services at lower costs than if they were provided individually) and density (the lowering of costs as larger flows can be channeled into each route), but suffers from potential congestion at the hub airports as traffic converges. From the airlines' revenue perspective, hub-and-spoke operations also allow them to enjoy economies of market presence   the ability to offer direct and indirect services between numerous airports.

These various network types overlap to a considerable extent with the types of commercial airlines that provide the bulk of services. The linear and radial services are features of low-cost or "no-frill" airlines such as Southwest in the United States and Ryanair in Europe. These carriers can standardize their fleets, fly from secondary airports, avoid scheduling difficulties by minimizing the number of connecting services they offer, and provide limited onboard and ground services because of the short duration of flights. In contrast, the "legacy," or full-service, carriers focus on channeling their traffic through hubs   for example, Delta Airlines through Atlanta and Northwest Airlines through Detroit   and deploy a mixed fleet to meet the needs of a more complex network, wide variations in flight lengths, and long-haul international services. To enhance the range and frequency of services, the legacy carriers often form global alliances. The Star Alliance, for example, includes United, Lufthansa, SAS, British Midlands, and other airlines. This allows for easier ticketing and fight connections as well as larger frequent flyer programs that reward passengers for loyalty to an airline or alliance.

Globally, the scheduled airline industry as a whole has not, at least since the 1980s, earned what would normally be seen as an economically viable return. Even in good years the operating margins of most airlines has been below the return offered by bank savings accounts. Excessive competition and, at least for a time, inflated costs account for this. This situation contrasts with most other elements in the air transportation value chain, with airports, air navigation systems, and global distribution systems all often making significant financial returns.

Besides scheduled airlines services, there are also charter services. These involve an entire aircraft's

capacity being sold usually to a tour operator who then sells these seats, often in conjunction with hotel rooms, rental cars, and the like, to leisure travelers. Until the 1990s, because their activities were less regulated than those of scheduled airlines, these operators were important in Europe, taking up to 20 percent of the air travel market. Regulatory reforms have reduced the importance of charter carriers and also their nature: many offer near scheduled services with regular flights, and seat-only sales, not requiring the addition of a hotel purchase, are common. There are also air taxis and business aircraft that offer personal air transportation, usually to business executives. In addition to civilian air services, the military in most countries operates large fleets of aircraft to move equipment and personnel, in addition to any direct combat-related activities, and these flights must be integrated into the overall air traffic control system.

Regulation Air transportation traditionally has been regulated very heavily. Economic regulation of fares, cargo rates, market access, and capacity became widespread after World War II, both internationally as countries tried to develop their own commercial air fleets, and domestically as air transportation became a mechanism for greater mobility and political cohesion within countries. Notions that airlines, airports, and air navigation services, if left unchecked, would become monopolies that would penalize users were used to justify economic regulation of fares and access. Social regulation was also widespread and often aimed at providing otherwise unremunerative services to remote regions as well as to ensure adequate safety.

Internationally, the UN International Civil Aviation Organization, established at the end of World War II at the Chicago Convention, regulates air transportation. The resultant agreements give sovereignty to countries regarding their own air space and set down the basis of negotiations between nations over international routes, called air service agreements. Until the late 20th century, most international traffic was severely controlled: often, only one carrier from each country could offer international services at a regulated fare, with a limited capacity, and with the revenues shared equally between

the two countries. Many domestic markets were also strictly controlled, and in some cases a state-owned airline was the only supplier of services at a regulated fare.

From the late 1970s on, there has been a global movement to introduce more market-based structures into the air transportation sector. This began when U.S. airlines removed rate and market entry controls (other than those retained for safety and security reasons) for domestic air cargo services in 1977 and from passenger services in 1978. The United States also initiated a large number of "Open Skies" agreements in the 1980s and 1990s that removed route access and fare controls from bilateral air service agreements with other countries but retained cabotage, the freedom of external airlines to operate within another country, and ownership controls. Parallel to this, by 1997 a gradual relaxation of international bilateral air service agreements within the European Union freed airlines services, both domestic and international, from economic regulation, including ownership rules within the EU.

The result of these changes has been an expansion in the number of commercial airlines and an increase in the diversity of services provided, not only by low-cost carriers on short-haul routes, but also by specialist airlines that offer only business class service on long-haul flights. In markets where regulatory reforms have been enacted, fares have been brought more closely into line with costs. At the same time, competition has lowered fares and forced airlines to become more efficient.

Airports The vast majority of the world's airports are either nationally owned or owned by a local (city or state) authority. The original reason for this was that airports, because of the large capital costs involved in establishing them, often are not commercially viable but nevertheless are important for strategic or local economic development reasons, and as a way to tie remote communities to larger cities. From the mid-1980s, airports have had to comply with strong regulations related to the noise, safety issues, and traffic congestion associated with them. Private sector involvement, however, has become more common under the state or local government

ownership umbrella, with airlines financing, leasing, or owning parts of airports (such as terminal buildings) or directly providing services (such as ticketing and ground handling). In some cases, airports have allowed private caterers, retailers, and the like to offer services on their sites.

Since the 1980s, many countries have moved to privatize entire airports, or to distance their ownership and operations from political processes. The British Airports Authority, which owns seven airports in the United Kingdom, including the three major airports serving London, was privatized with equity capital. Other airports have been bought out as going concerns by commercial companies with no equity holdings. Privatization, however, often brings with it challenges of regulating monopoly power; most cities only have one airport, or at most two. The approach pursued in many cases involves "price-capping," with a regulator controlling the ways in which prices charged by airports for take-offs and landings may change.

In South America and some parts of Africa, various forms of concessions have allowed states to retain their ownership of an airport but hand their entire management and investment strategy over to private companies for long periods. Such approaches are aimed at bringing private capital and expertise into airport development where public finance is limited and the local market does not have an adequate supply of skilled workers and managers.

Air Navigation Systems Providers Air transportation involves the movement of aircraft along corridors akin to three-dimensional railway tracks. In some cases the traffic flows freely with air navigation system providers (ANSPs) supplying navigation and weather advice. Most commercial aviation, especially in congested air space, is subjected to air traffic control to prevent accidents. These involve en route controls at higher altitudes and tower controls around airports.

Reforms in air transportation infrastructure have been slower to materialize than for airlines. Most ANSPs traditionally have been state owned but in recent years a number have been privatized or "corporatized" (turned into nonprofit entities either in-

dependent or government-owned but not managed). For example, NAV CANADA became a private not-for-profit corporation in 1996; Airservices Australia became a government corporation in 1988; and the UK air transportation system, NATS, became a public-private partnership in 2001. France retains state ownership of its ANSP but allows Direction des Services de la Navigation Aérienne access to private financial markets and to levy user fees. The Federal Aviation Administration in the United States outsources some activities but is financed by taxes. Because of the increase in international air transportation and the need for greater integration of systems, some countries are developing coordinated strategies for delivering air navigation systems (ANSs). Within Europe, for example, EUROCONTROL has the remit for creating a Single European Sky that ultimately links the various national ANSs.

The changes have come about as new ways are being sought to finance the modernization of facilities and to improve their efficiency. ANSs are highly capital intensive and are continually being improved, but adopting the new technology is expensive and often has to be tailored to the peculiarities of the existing system rather than using generic hard- and software.

One of the main difficulties in enhancing air traffic management performance through regulatory reform has been the concern with aviation safety. Academic studies show a greater aversion to being killed in a plane crash, for example, than in an automobile accident. The evidence to date is that, although there are some geographical areas where the record is poor, the overall safety of air transportation has continually improved over the years in terms of both mortalities and morbidity per passenger mile flown.

Overall, air transportation is the fastest-growing transportation mode in the world for both passenger and freight traffic. It has traditionally been highly regulated but since the late 1970s, the removal of economic regulation from many markets has led to improved efficiency, innovation, and lower costs and fares, without any reduction in safety standards. There is now diversity in the types of airlines that

offer services and also in the forms of infrastructure that is required to support flights.

See also shipping; trade in services

FURTHER READING

Airline Business. Monthly. Offers articles on the state of the air transportation industry and regularly provides detailed information at the global level on issues such as the membership of airline alliances, the scale and operational features of airports, and safety matters.

Boeing Commercial Airplane Group. Annual. *Current Market Outlook*. Provides reviews of the state of the air transportation market and long term forecasts of traffic levels.

Doganis, R. 2006. *The Airline Business*. 2d ed. London: Routledge. Explains how the airline market works and the issues that confront both operators and policy makers.

Journal of Air Transport Management. Six times a year. Provides analysis at an academic and professional level of all aspects of air transportation.

KENNETH BUTTON

■ **Andean Community**

The Andean Community (CAN), as this organization is currently known, is a regional integration agreement among Colombia, Ecuador, Peru, and Bolivia. The community was originally known as the Andean Pact and was created in 1969 to reverse the stagnation of the Latin American Association of Free Trade and address the integration and development needs of the Andean countries (Venezuela, Colombia, Chile, Ecuador, Peru, and Bolivia). Venezuela became part of the Andean Pact in 1973, and Chile withdrew from the pact in 1976 to pursue more liberal trade policies. Initially, the pact sought to harmonize policies, define a common external tariff, liberalize intraregional trade, regulate foreign direct investment in the region, and organize production across member Andean countries by encouraging the development of promising industries. Later on, the pact supported the agenda of becoming part of wider economic agreements such as Mercosur a

regional economic agreement among Argentina, Brazil, Paraguay, and Uruguay and the Free Trade Area of the Americas (FTAA) to be consistent with the General Agreement on Tariffs and Trade and the World Trade Organization principles.

The initial strategy of the Andean Pact was based on the import substitution, or closed regionalism, model that predominated in Latin America during the 1970s. According to this model, the government must coordinate economic policies and regional development plans in order to direct production toward the intraregional market. The consequence of this model is that protected rent activities (activities that generate rent because of government protection through tariffs or subsidies) develop, mainly in the industrial sector, which are financed in part by the revenues generated by primary-resource-intensive exports (agriculture, mining, and energy). Intraregional trade only increased from 1.7 percent of total exports in 1970 to 4.5 percent in 1979. This early stage of the Andean Pact failed for several reasons: many products were exempted from the tariff liberalization process; a clear consensus about the common external tariff was lacking due to significant differences in the level of protection of each Andean country; the production requirements established by the Andean Pact did not match the trade needs of each country, especially after the foreign debt crisis; the market was too small; and trade activity was directed mainly to the members of the Andean Pact. Therefore, the Andean countries were limited in their capacity to generate new foreign exchange, which became very important for paying the increasing foreign debt (Edwards 1993). The lack of coordination of macroeconomic policies led to exchange rate imbalances and to differences of protection among the Andean countries. (In 1980, about 25 percent of items included initially in the tariff list were exempted.)

These macroeconomic imbalances partially generated by the closed regionalism model contributed to the foreign debt crisis that exploded in the early 1980s. The adjustment policies applied to solve the crisis led to a contraction of the trade preferences among the Andean countries, thereby reducing the

trade during the mid-1980s. By 1985 the Andean Pact was practically moribund. Intraregional trade did not follow the initial industrial planning, and only about a third of the investment programs (machine tools, petrochemical, and automobile sectors) were approved. Nevertheless, the Andean Pact was revived with the Quito Protocol, which was signed in 1987 and later modified over the course of several presidential meetings. The most important modification, the Trujillo Protocol of 1996, resulted in the name change from the Andean Pact to the Andean Community of Nations, a new structural organization, and a shift in emphasis from closed regionalism (inward integration) to open regionalism (outward integration with the rest of the world) (Reynolds 1997; and ECLAC 1994). The establishment of the Andean Free Trade Zone (AFTZ) in 1993, and the Andean customs union in the form of an Andean common external tariff in 1995, spurred private initiatives and innovative rent-seeking activities (instead of protected rent activities) aimed at achieving an efficient allocation of resources and exploiting the competitive advantages of the region. This increased efficiency and innovation is the main reason behind the shift toward the open regionalism model. The AFTZ was completed in 2006 with the full incorporation of Peru.

Using CAN as an example, Creamer (2003) demonstrated that economic integration by stages into wider regional agreements may lead to an improvement of intraregional trade and total trade, and not to a contraction of extraregional trade as during the period 1980 97. Taking a simple indicator, the trade balance of CAN exports minus imports with the rest of the world increased from U.S. $1.196 trillion in 1969 to U.S. $31.435 trillion in 2005 (Andean Community 2006). However, this growth is mostly explained by the significant improvement of the trade balance of Venezuela. Venezuela accounts for 90 percent of CAN's trade surplus with the rest of the world. This trade balance may be substantially affected because of the withdrawal of Venezuela from CAN in 2006, although this effect may be partially compensated by the integration of CAN with Mercosur and the incorporation of Chile. The Andean Council of Foreign Affairs Ministers and the Andean Community Commission accepted Chile as an associate member in September 2006.

Governance During the 1990s, CAN changed its emphasis from a trade-oriented agreement into a political, social, and economic integration agreement in the spirit of the European Union. This transformation required the creation of the Andean Presidential Council and the conversion of the Cartagena Agreement into a General Secretariat in 1997. The set of all institutions that support the mission of CAN is the Andean Integration System (SAI). The political institutions that are part of SAI and have representatives of each member country are the Andean Presidential Council, composed of the presidents of member countries, which defines the strategic priorities of CAN; the Andean Council of Foreign Affairs Ministers, organized by the foreign affairs ministers of member countries, which is responsible for the foreign policy of CAN, agreements with third parties, and the election of the general secretary; the Andean Community Commission, in coordination with the Andean Council of Foreign Affairs Ministers, which is responsible for the intraregional policies of CAN; the Andean Community General Secretariat, which is the executive body of CAN; the Andean Community Court of Justice, which is the judicial body that interprets the Andean Community laws and solves internal disputes; the Andean Parliament, the legislative arm of CAN, which harmonizes member countries' laws; and the Business and Labor Consultative councils, which represent business and labor organizations, respectively, and advise the Andean Council of Foreign Affairs Ministers. Additionally, the Andean Development Corporation, which is responsible for promoting trade, investment, and economic growth in the region, and the Latin American Reserve Fund, which provides funds to the member countries to correct short-term macroeconomic imbalances and coordinates monetary and fiscal policies, also are part of the SAI. Finally, institutions of SAI responsible

for social policies are the Simón Rodríguez Agreement, which coordinates social and labor policies; the Andean Health Organization Hipólito Unanue Agreement, which coordinates health policies; and the Simón Bolívar Andean University, which promotes academic activities that are relevant to the integration and development of CAN.

Most of CAN's political institutions have not had a major impact on the internal policies of the Andean countries. These countries are still very much under the influence of the major financial multilateral organizations. Hence, CAN's political integration has been very limited. The Andean Development Corporation, however, has become a major source of funds to support the development policy of CAN. This organization has also had an impact in social areas, providing funds for projects that generate employment, support microentrepreneurs, and improve the productivity of the region. The Simón Rodríguez Agreement has helped to integrate educational systems in the Andean countries, and the Simón Bolívar Andean University has been invaluable in supporting the professional training of social scientists. However, major regional challenges include the development of human capital in technical areas and a significant increase in research and development investment to attract foreign capital and outsourcing opportunities to the region.

Andean Customs Union and Free Trade Zone
The Andean customs union, as it has functioned since 1995, establishes four basic tariff levels: 5 percent for raw material and industrial output, 10 percent for intermediate output, 15 percent for capital goods, and 20 percent for final goods. There are some exceptions to this common external tariff. For agricultural products, price bands help protect Andean agricultural products from subsidies and price variations in the international market. In 1997, Peru decided to join the Andean customs union and started a program of tariff reduction with Columbia and Ecuador. The Andean common external tariff covers about 90 percent of imports. The Declaration of Santa Cruz of the Andean Community, signed in

January 2002, introduced a new structure of the common external tariff that includes Peru. However, the new common external tariff had not been implemented and was still under review by a high level advisory group since August 2007.

CAN represented the Andean countries in the FTAA negotiations and lobbied the U.S. government for the extension of the Andean Trade Preference Act of 1991 into the Andean Trade Promotion and Drug Eradication Act of 2002. Also, CAN renewed the Generalized System of Preferences with the European Union for the period 2006 15 and at the time of this writing was negotiating a trade association agreement. In April 1998, CAN signed an agreement with Mercosur to create a free trade zone. As a result of this agreement, 80 percent of the trade between Mercosur and CAN was freed in January 2005. The remaining 20 percent will be liberated during the next 14 years. Additionally, the members of Mercosur became associate members of CAN, and Mercosur reciprocally conferred associate membership on the members of CAN.

The withdrawal of Venezuela in 2006, the potential signing of free trade agreements between the United States and Columbia, Ecuador, and Peru, and the formation of a South American free trade zone as agreed during the South American presidential meeting of 2004 are signs of the difficulties facing CAN. As long as CAN maintains its own identity, free trade agreements with the United States or with the other South American countries may indicate the success of the open regionalism process that has characterized the Andean Group since the 1990s. In fact, a South American free trade zone will open a market of 377 million inhabitants with a gross domestic product of U.S. $1.493 trillion and exports of U.S. $305.3 billion for 2005.

CAN offers significant advantages to its members, especially once the Andean Common Market is established. This market will enable the free movement of goods, services, capital, and people. CAN envisions that, in this way, it gradually will integrate itself into the world market, either through its participation as a subregion in a South American free trade

zone or in a free trade agreement with the United States.

See also customs unions; free trade area; Free Trade Area of the Americas (FTAA); import substitution industrialization; Mercosur; regionalism

FURTHER READING

Andean Community. 2006. "37 Años de Integración Co mercial, 1969 2005." http://www.comunidadandina .org. A compendium of the relevant trade statistics of the Andean Community since its foundation.

Creamer, Germán. 2003. "Open Regionalism in the An dean Community: A Trade Flow Analysis." *World Trade Review* 2 (1): 101 18. An analysis of Andean Commu nity trade that demonstrates a pattern of open region alism.

ECLAC (Economic Commission for Latin America and the Caribbean). 1994. *Open Regionalism in Latin America and the Caribbean.* Santiago, Chile: ECLAC. A policy analysis of the Latin American integration process from the open regionalism perspective.

. 2006. *Panorama de la Inserción Internacional de America Latina y el Caribe. Tendencias 2005 6.* Santiago, Chile: ECLAC. Annual publication about Latin Amer ican international trade with a chapter on Latin Ameri can economic agreements such as the Andean Com munity.

Edwards, Sebastian. 1993. "Latin American Economic In tegration: A New Perspective on an Old Dream." *The World Economy* 16 (3): 317 39. A macroeconomic ex planation of the limited success of Latin American eco nomic agreements.

Reynolds, Clark W. 1997. "Open Regionalism: Lessons from Latin America for East Asia." Kellogg Institute Working Paper No. 241. Notre Dame: University of Notre Dame. Available at http://kellogg.nd.edu/publi cations/workingpapers/WPS/241.pdf. Presentation of the open regionalism experience in Latin America and East Asia. This working paper is based on Clark W. Reynolds, Francisco Thoumi, and Reinhart Wett mann's "A Case for Open Regionalism in the Andes: Policy Implications of Andean Integration in a Period of Hemispheric Liberalization and Structural Adjustment" (1994), a typescript report prepared for the Friedrich Ebert Stiftung Foundation and USAID. North America

Forum. Palo Alto: Stanford University. This is the first research that applies the term *open regionalism* to the Andean integration process and shows how the Andean Community can change its emphasis from inward in tegration to outward integration with the rest of the world.

GERMÁN CREAMER

■ anti-dumping

Many countries reserve the right to impose import taxes on foreign products that have been found to be "dumped" into their domestic economies (that is, sold at less than their "normal" prices). These tariffs, known as anti-dumping duties, have become a crit ical but controversial part of the multilateral trading system that has developed first under the General Agreement on Tariffs and Trade (GATT) and then under its successor, the World Trade Organization (WTO). Anti-dumping duties are one of the few ways consistent with international obligations by which governments may increase tariffs beyond levels negotiated in international trade talks.

WTO rules allow its members to impose these duties if the individual governments determine that foreign firms "dump" a product that damages a do mestic industry producing a similar product. In common usage, anti-dumping duties are used to counteract "unfair" import pricing practices by for eign competitors.

Dumping is defined in two principal ways within the WTO system. The first is when a foreign firm sells in a domestic market at a price below the price it sets for the same good in its home market, known as "price-based" dumping. In others words, interna tional price discrimination is an actionable practice under multilateral trade rules. The second basic definition, known as "cost-based" dumping, in volves a foreign firm selling its product in the do mestic market below its cost of production. In practice, this usually means that foreign firms are selling below the average total cost of producing the item. Anti-dumping duties may be imposed under WTO only if the dumping margin exceeds a de

minimis level of 2 percent (i.e. the percentage difference between the "normal" value and the price charged in the export market); dumping below this level is not subject to duties.

It is important to note that dumping is based on the pricing practices of individual firms in specific countries. Consequently, anti-dumping duties may vary across firms within a particular exporting country. This means that customs officials in a country that has imposed anti-dumping duties must pay very close attention to which particular foreign firm is exporting the product and from which country. This can impose important administrative burdens and in principle may provide incentives for foreign firms to try to circumvent the duties by mislabeling the product's country of origin.

Anti-dumping duties are a recognized exception to a number of core WTO principles. For example, the "most-favored-nation" (MFN) principle means that a member commits to impose the same tariff on the same product for all WTO member countries. Anti-dumping duties, in contrast, can vary across countries for the same product. WTO rules also normally require governments not to raise duties beyond the tariff levels agreed to through multilateral negotiations. Members are freed from this obligation, however, if foreign firms are found to be dumping. Finally, WTO rules require "national treatment" so that foreign firms and domestic firms are treated in similar ways. Under anti-dumping rules, foreign firms selling below average total cost (i.e., total revenue is below total cost) may be subject to sanction while a domestic firm that does not cover its costs will not face similar fines. The fact that anti-dumping is a WTO-allowed exception to these critical principles of the multilateral trading system suggests the importance that member governments place on discouraging this type of foreign pricing behavior.

Administration of Anti-dumping Rules Anti-dumping investigations are usually initiated by domestic firms concerned about foreign competition. The importing country's government then must undertake two separate investigations, one to determine whether there is evidence of dumping and the other to find out whether the imported products are harming the domestic industry. The former requires information about foreign firm pricing and the latter data on the competing domestic firms' economic conditions.

A domestic administering agency responsible for calculating dumping margins requires different detailed information about foreign firm behavior. For example, it must collect pricing data in the foreign firm's home market if it uses price-based methods. If it uses cost-based methods, it must obtain information about the foreign firm's production costs. Both methods require foreign firm cooperation if precise dumping margins are to be calculated. Some foreign firms may be reluctant to open their books to another country's investigating agencies and may decide not to cooperate. WTO rules allow for an administering agency to use other sources of information, including allegations of the domestic import-competing firms, to complete the investigation. The threat of using domestic allegations serves as an inducement to the foreign firm to cooperate by providing needed data.

Calculating dumping margins can also be complicated if there is pervasive involvement of the government in the economic life of a country. For example, before the demise of the Soviet Union, an agency investigating a charge of dumping by a Soviet firm would not be able to use any information based on administratively determined prices. The multilateral trade system consequently recognizes different anti-dumping procedures in cases of "nonmarket economies" (with the People's Republic of China as the most prominent current example). In order to conduct an investigation involving such an economy the domestic agency may collect information about the quantities of inputs used by the Chinese firm. The agency will then apply the input prices of a third country deemed to be at a similar stage of development (e.g., Bangladesh or India) to calculate a proxy cost of production. This can then be used to determine the dumping margin by comparing the constructed value with the price charged in the export market.

The second part of an anti-dumping investigation requires that a domestic agency determine whether

the dumped import causes "material injury" to a domestic industry producing the "like product." This requires that the domestic industry provide critical information about its current condition. Various indicators of material injury are used, including declines in domestic sales, employment, and profits, or the degree to which the dumped import prices are below domestic prices. In addition, the "threat" of material injury can also be a basis of imposing anti-dumping duties. Such a threat might be found if a firm found to be dumping has recently added significant export capacity.

According to WTO rules, anti-dumping duties finally imposed can be calculated in two distinct ways. One is that the import duties are equal to the dumping margins (i.e., difference between the "normal value" and the export price). This procedure is used in many jurisdictions, most notably in the United States. Those favoring this approach argue that anti-dumping duties are allowed because dumping itself is the ultimate problem, so the policy response must eliminate any dumping. The other approach is that the anti-dumping duties can be limited to the lesser of the dumping margin or the minimum duty required to eliminate injury to the domestic industry. This second version, the so-called "lesser-duty rule," is used in the European Union. Supporters of the latter approach often argue that the main goal of the anti-dumping process is to eliminate injury, not dumping itself.

WTO rules also allow alternatives to increased duties in anti-dumping cases. "Price undertakings" involve a settlement between the foreign firms accused of dumping and the importing country government. Typically, foreign firms offer to raise their export prices to eliminate any dumping in order to avoid duties placed on their products. If the investigating domestic government agrees to the offer, it will suspend the anti-dumping investigation as long as the foreign firms keep their exports prices sufficiently high. Price undertakings are particularly common in the European Union.

Anti-dumping orders may be in place for long periods. There is no specific time frame for their removal as there is with safeguard measures, which have an eight-year maximum. WTO members did agree in the Uruguay Round to require governments to review anti-dumping orders every five years after the original imposition of duties. In particular, duties must be taken off unless an investigation determines that their removal likely would lead to lead to a resumption of dumping and material injury to the domestic industry.

Anti-dumping actions, like other aspects of the multilateral trading system, are subject to the WTO review. This means that member governments can ask the WTO dispute settlement bodies to rule whether countries imposing anti-dumping duties have lived up to their obligations. This setup allows WTO signatories to develop and use their own systems and procedures to protect their domestic industries against imports found to be dumped and causing injury but allows exporting countries some recourse if they believe that the anti-dumping duties are unjustified or implemented inappropriately.

WTO rules give governments the right to impose duties on injurious dumped imports but do not require it. WTO obligations allow governments to impose anti-dumping tariffs only if doing so would be in the "public interest." This provision reflects the fact that import taxes raise the price paid by domestic consumers, including industrial users that import intermediate inputs. For example, the European Union's procedures require that national representatives evaluate the impact of anti-dumping duties on broader EU interests. In sharp contrast, other countries such as the United States prohibit administrators from considering consumer interests when making the decision of whether to impose anti-dumping duties.

Anti-dumping procedures are similar to another WTO-consistent method to impose import restrictions. Countervailing duties (CVDs) are allowed by WTO rules if exports have been subsidized by a foreign government and those exports cause material injury to a domestic industry. Such subsidies are deemed to be unfair to domestic competitors since it involves competition with a government, not just with a private company. Thus, CVDs procedures have a structure similar to anti-dumping duty sys-

tems (i.e., "unfair" competition plus material injury) but involve foreign government actions rather than private firm pricing decisions as in the case of anti-dumping.

Supporters and Critics Anti-dumping duties are a controversial part of the WTO. Many supporters argue that anti-dumping is a necessary response to unfair trade. Critics regard it as simple protection with a rhetorical flourish.

Many of those who support anti-dumping's role in the international trading system argue that it counteracts foreign predatory pricing practices. *Predatory pricing* refers to a strategy under which a firm charges very low prices in order to drive out competitors so that it can later increase prices after achieving a monopoly position. Domestic antitrust and competition policy are at least partially built on these ideas. In the case of international competition, supporters of anti-dumping argue that international price discrimination or pricing below average total cost is akin to predatory pricing by foreign firms.

Supporters also argue that anti-dumping duties are necessary to counteract foreign firms operating from a "sanctuary market." This refers to the possibility that formal and informal trade barriers allow foreign firms to receive higher-than-normal profits in their own market, which they then use to offset losses in the export market as they try to drive out domestic competition. This also is sometimes seen as a part of a predatory pricing strategy.

Anti-dumping is often criticized by those favoring trade liberalization. One of the most important criticisms is that anti-dumping is a just another form of protectionism since it involves an increase in duties on imports. In this view, anti-dumping duties reduce access of domestic consumers to foreign goods and result in greater damage to consumers than any benefits accruing to the domestic producing industry. Critics also note that anti-dumping investigations are extremely complicated, especially regarding the calculation of dumping margins. Domestic agencies must make a myriad of decisions about how to treat transactions of an individual foreign firm and how to calculate foreign costs of production. Uncertainty about how the investigating agency will calculate those margins means that it is difficult for foreign firms to know what margin they will finally face. One reaction by foreign firms is to raise their prices even in the absence of an anti-dumping allegation; the mere presence of an anti-dumping system may result in higher prices for consumers. Critics also point out that there is no requirement to determine whether or not there is any evidence of a sanctuary market or predatory intent by foreign firms before imposing anti-dumping duties. Finally, they note that domestic firms that sell their goods for different prices in different cities domestically (i.e., engage in domestic price discrimination) or sell domestically below average total cost (i.e., losing money on domestic sales) are not subject to sanction by the domestic government. In other words, anti-dumping duties punish foreign firms for undertaking practices acceptable if done by domestic firms.

Some observers, even among those who doubt whether international predatory pricing is an important threat, believe that the presence of anti-dumping rules may allow politicians to engage in broader trade liberalization than would otherwise be possible. This view holds that anti-dumping rules can act as a kind of safety valve; governments may be able to deflect pressures for broader protectionist measures by giving domestic industries an administrative process under which they can petition for import restrictions. This could allow governments to raise duties on only a small number of narrowly defined products even as restrictions are lifted on a whole array of other imports.

Anti-dumping in Practice Anti-dumping has a long history, especially in the English-speaking world, but it has spread across much of the world in recent years. The world's first anti-dumping law came into effect in Canada in 1904, followed shortly by New Zealand (1905) and Australia (1906). The United States implemented its first anti-dumping law in 1916. The countries that would eventually make up the European Union were also early users of anti-dumping provisions. This group of countries constitutes the traditional users of anti-dumping; up until 1986, worldwide use of anti-dumping was almost exclusively restricted to this group. Beginning

in the mid-1980s, countries such as Mexico, Argentina, and Brazil began to use the procedure more frequently. After the establishment of the World Trade Organization in 1995, many other countries began to use anti-dumping, including most notably India, the Republic of Korea, South Africa, and Turkey. Japan, though frequently a target of anti-dumping actions by importing countries, has only rarely used anti-dumping to restrict imports. The People's Republic of China has used anti-dumping more frequently in the post-2002 period but that country's involvement principally has been as the most frequent target of anti-dumping actions.

Anti-dumping duties have been imposed on a wide variety of manufactured, agricultural, and commodity products. The steel and chemical industries have been the most frequent users of the anti-dumping process. For the period between 1980 and 2002, approximately 50 percent of all anti-dumping petitions involved these two industries. This may reflect the very high fixed costs of both industries, which make them vulnerable to downturns in their industries and create incentives to expand production through exports to try to lower average production costs. The steel industry has been particularly active in the United States in supporting the role of anti-dumping in the multilateral trading system.

In short, the anti-dumping system receives strong support among some important members of the World Trade Organization but is controversial among nations especially reliant on exports.

See also countervailing duties; nondiscrimination; nontariff measures; World Trade Organization

FURTHER READING

Blonigen, Bruce, and Thomas Prusa. 2001. "Antidumping." In *Handbook of International Trade*, vol. 1, edited by Kwan Choi and James Hartigan. Oxford: Blackwell, 251 84. Provides a literature review of academic research on anti dumping.

Finger, J. Michael, ed. 1993. *Antidumping: How It Works and Who Gets Hurt*. Ann Arbor: University of Michigan Press. Details administrative practices of agencies responsible for anti dumping investigations.

Mastel, Greg. 1998. *Antidumping Laws and the U.S. Economy*. Amonk, NY: M. E. Sharpe. Analyzes some of the major arguments in favor of the use of anti dumping laws in the world trading system.

MICHAEL O. MOORE

■ anti-globalization

The spectacular growth in the intensity, scope, and visibility of globalization (understood here as the increasing interconnectedness of individuals, groups, companies, and countries) since 1990 has been accompanied by a parallel growth in anti-globalization a broad term used to characterize a public debate over the shaping, slowing, or rejecting of globalization.

Driven by the growth of international economic integration and international institutional arrangements and the spread of ever-denser networks of global communications, a debate has arisen since the 1990s around concerns regarding the distributional benefits of globalization, the desirability and impact of different types of policy, and the nature and representativeness of the political institutions that decide on global policy issues. The public face of this debate has been notable for the high-profile role of nonstate actors and for its focus on perceived shortcomings in current systems of global governance.

The terms most commonly used for this global debate, particularly with the involvement of nonstate actors *anti-globalization* and *the anti-globalization movement* were used most frequently between the failed World Trade Organization (WTO) Ministerial in Seattle in 1999 and the Group of Eight (G8) Summit held in Genoa in 2001 amidst violence by both police and demonstrators. These terms are misleading, as participants in this debate are neither solidly "anti-globalization," nor a single movement. Under the umbrella of these labels is a wide variety of actors with often sharply differing philosophies, objectives, and assumptions. Yet since the late 1990s they have proved a useful if contested and in large part inaccurate shorthand to describe what many

identify as a new and important force in global politics.

In order to understand the nature of the current debate about globalization as well as what might happen next to the "anti-globalization movement" and its actors there is a need for disaggregation and a reflection on the context in which it exists.

Origins Although its political and social origins are diverse, anti-globalization in its broadest sense can be seen as a response to the economic and political events of the period since the early 1970s and their most visible symbols, the institutions of global economic governance. In the North, the oil crisis and the suspension of dollar convertibility in 1971 marked the end of the "long boom" of post-1945 Keynesianism. They also triggered the meteoric rise of global capital markets, which made earning and keeping "market confidence" an increasingly important determinant of government policies. In the South, the Mexican government's near-default on its foreign debt in 1982 marked the end of the postwar era of import-substituting industrialization and began a long and painful period for developing countries, characterized by the burden of massive foreign indebtedness and the rise in political influence of the International Monetary Fund (IMF), the World Bank, and international capital markets, all three of which ushered policymakers away from development policies focused on the domestic market and toward a strategy of export-led growth.

These developments helped drive the rapid expansion of trade and investment flows, as large parts of Latin America and Asia adopted export-led growth strategies, and the countries of the former Soviet empire were rapidly, if partially, absorbed into an increasingly integrated global economy. The term *globalization* quickly became the shorthand for this model of expansion a heady and complex mix of technological, economic, political, and cultural change.

Globalization was accompanied and underpinned by a set of interlocking institutional developments at international and national levels. First, the existing structures of global economic governance were overhauled. The World Bank and the IMF redefined their roles, moving swiftly away from Keynesian operating principles to become bastions of neoliberalism. A web of bilateral, regional, and global international trade and investment agreements, culminating in the creation of the WTO in 1995, bound the new system in place. These economic and political trends unified in opposition a diverse array of actors, however. Downsizing and corporate restructuring, privatization, the erosion of workers' rights, and the changing nature of production and supply chains activated opposition from the labor movement in both the North and the South. Global warming, unsustainable growth, and the depletion of resources created hostility from environmentalists, who were further outraged over the perceived threat to environmental legislation from trade rules in the WTO for example, when four Asian nations successfully challenged provisions of the U.S. Endangered Species Act forbidding the sale in the United States of prawns caught in ways that kill endangered sea turtles. The erosion of the nation-state and of democratic institutions antagonized proponents of state-led development, democrats, and some on the political right. Increasing corporate power and social inequality galvanized the traditional left and a whole host of other left-of-center actors. Structural adjustment programs and growing Southern marginalization and inequality radicalized civil society (the term for nonstate civic and social organizations) and some political parties in the developing world.

The changed international institutional landscape also aided the growth of this opposition. This shift of power to international institutions and the growth in the range and reach of their activities were not well linked into the traditional accountability mechanisms of states. These inadequacies in global governance have raised the profile of attempts by nonstate actors to make these institutions more accountable. The increasing global reach of the World Bank and the IMF has provided common rallying points for protest, and the founding of the WTO in 1995 in particular put an institutional face on what had

previously been an amorphous process   a gift to the protest movement.

Who's Who? At first glance, anti-globalization seems an incongruous political mix of contradictions, colors, and cultures, in part vocal and aggressive, in part quiet and conciliatory. Although it defies firm categorizations, the movement can be roughly divided into three strands: statists, alternatives, and reformists.

The statists believe the current process of globalization has been a disaster and seek to defend and rebuild the role of the state in economic management after the neoliberal assault that began in the 1980s. This group is dominated by the traditional left, some sections of the labor movement, and a large proportion of Southern activists. Through this group runs a strong sense of rejectionism and even conservatism. Some, such as a few of the U.S. labor unions protesting at the WTO Ministerial in Seattle, want to retain the state's ability to protect domestic industries from cheap imports. Others, such as the prominent Filipino activist Walden Bello, reject the terms of globalization outright, feeling that any alternative, including the abolition of the IMF and WTO, could not fail to be an improvement on present realities. Despite its focus on the nation-state, this group retains a strong sense of internationalism.

The alternatives are both highly visible and the hardest to define, though often labeled "anarchist." This element of the movement is strongly driven by cultural concerns and best understood in cultural terms. Its members   be they ecologists running organic businesses, followers of the *Small Is Beautiful* author E. F. Schumacher, activists seeking to "deconstruct" corporate power and global brands, or Zapatistas who wish to gain rights and land and make a statement about globalization's marginalizing effects   reject globalization in passing but concentrate more on building small-scale alternatives. These groups oppose the encroachment of the market or the market's power relations on their cultural or political spaces. Most are also small, decentralized, and strongly anticorporate.

The reformists make up the majority of formally structured groups involved in the movement, or at

least dominate the thinking of the movement's leadership. Their aim is partial change to offset current injustices and inequalities. The reformists act within current political systems and advocate gradualism and peaceful change. Most accept a role for the market but believe it must be better regulated and managed in order to achieve socially just and sustainable outcomes. This group includes some trade unions, faith groups, charities and development organizations (such as Oxfam), and most mainstream environmental groups (including Friends of the Earth), as well as issue-specific campaigns such as Drop the Debt and the Tobin tax.

The reformist current has also made strong inroads into global and national politics, going far beyond "the usual suspects." The *Financial Times*, Gordon Brown, Nobel Prize winning economists such as Amartya Sen and Joseph Stiglitz, Kofi Annan, the corporate social responsibility movement, Jeffrey Sachs, and George Soros could all be called reformists. As the economist Meghnad Desai puts it, "The reformists view themselves as the only true defenders of globalization. They believe that both isolationist calls to reverse the process and supporters' insistence on 'ultra-liberal' forms of global capitalism are bound to de-rail globalization, with tragic consequences" (Desai and Said 2001, 68).

This attempt to disaggregate the movement warrants several caveats, however. Many nongovernmental organizations (NGOs) and even individuals span more than one current: for example, Friends of the Earth is both reformist and alternative. Author Naomi Klein, one of the movement's most prominent figures since the publication of her book *No Logo*, may base her critique of globalization primarily in cultural terms and is a source of inspiration to the anticorporate wing of the movement, but is herself essentially a progressive reformist. Within mainstream NGOs, supporters and Southern partners often espouse more radical options than the full-time staff and leaders.

Nor does this picture do justice to the depth and breadth of the movement in the South. The largest protests against the WTO have been in India. Brazil is rapidly becoming a center of the movement, as

evidenced by the huge gatherings of activists in Porto Alegre since January 2001 as part of the World Social Forum, held as a "people's response" to the World Economic Forum business summits in Davos. The movement in the North draws inspiration and guidance from a number of prominent Southern intellectuals such as Vandana Shiva (India), Martin Kohr (Malaysia), and Walden Bello (Philippines, but based in Bangkok) and the work of the NGOs to which they belong. Finally, none of these categories describes the nihilist currents, few in number in Seattle but significant at the G8 summit in Genoa in 2001, which used the protests as a platform for street violence rather than political debate.

Global institutions also play a key unifying role, both in providing focus to the movement and in creating the backdrop against which the movement has thrived. The movement is seen by some to be an important international player in its own right, helping to redefine public notions of democracy, accountability, and collective mobilization. The *Financial Times* identifies it as a "fifth estate," a valuable global counterbalance in a world of aging and often inadequate global institutions. The movement's reformist currents in particular have played an increasing role in what some academics have called "postsovereign governance."

But attitudes toward global institutions also mark a significant cleavage in the anti-globalization movement. For reformists, engagement is vital if change is to be achieved. Engagement has provided a small measure of greater transparency, participation, and popular pressure as NGOs enter policymaking channels and new mechanisms are created in an attempt to bridge gaps within highly imperfect existing structures. For rejectionists, by contrast, global institutions (in particular the WTO, IMF, and World Bank) are fundamentally illegitimate and unreformable, to be abolished rather than improved.

Main Concerns In general, the concerns of NGOs and civil society stem from an assertion that although globalization has led to benefits for some, it has not led to benefits for all. The benefits appear to have gone to those who already have the most, while many of the poorest have failed to benefit fully and some have even been made poorer. A linked concern of NGOs is that the drive for liberalization is based too much on dogma and ideology rather than on careful examination of the evidence and assessment of likely impact.

Equity and redistribution are seen as the missing link between globalization and poverty reduction. NGOs argue that improved equity within states leads not only to faster poverty reduction for a given amount of growth but also to faster growth. What is good for poor people is good for the economy as a whole. Yet up to now, globalization is seen to be frequently linked to increasing inequality, at both the national and the international levels. NGOs also argue for redistribution of wealth between developed and developing countries through debt relief and increased aid flows.

NGOs highlight research that points to the importance of national differences. The same policy reforms have different outcomes in different countries, depending on the structure of the economy, the initial distribution of assets, and the nature of economic and political institutions. Policy responses to globalization should be appropriate to particular cases in terms of the instruments used, the sequencing of reforms, and the combination of policies implemented.

Even though the evidence points to the importance of diversity, however, developing country governments are pushed by international rule-making, whether under the auspices of the WTO, through the pressures exerted by structural adjustment packages, or by the need to reassure the markets, toward greater homogeneity of policy response. The challenge for policymakers is to find ways of ensuring that national and international rule-making accommodate appropriate diversity of policy rather than reduce diversity to a minimum.

One of the lessons of recent years is held to be that liberalization and deregulation have very different costs and benefits when applied to the three areas of financial flows, direct investment, and trade. There has been concern that the frequency and severity of financial crises in recent years demonstrate the need for serious reforms of the global financial

architecture. Crises are seen to hurt the poor disproportionately and increase inequality, making the achievement of growth favorable to the poor harder thereafter.

One of the most high-profile areas of public concern (demonstrated by the impact and worldwide sales of Naomi Klein's *No Logo*) is that the increasing size and dominance of transnational corporations is making them more influential and less accountable. Public concern over excessive corporate power has led to calls for increased international regulation and has put pressure on companies to regulate themselves through the introduction of codes of conduct for themselves and their suppliers. In financial circles, this pressure has been accompanied by a greater awareness that successful companies must take into account a range of nonfinancial risks, including social, environmental, and ethical issues.

There are also fears that competition between countries wishing to attract foreign investment and technology could lead to a "race to the bottom" in terms of tax incentives and labor market suppression, thereby minimizing the potential social benefits offered by the private sector. Critics argue that the impact of foreign direct investment on employment, export performance, and domestic industry is not guaranteed, and that governments must be able to provide a regulatory framework to maximize the benefits and minimize the costs.

Finally, although most mainstream NGOs believe strongly that it is essential to have rules governing international trade, they severely criticize the particular set of rules established in the WTO. They see a multilateral trading system as necessary to ensure that weaker nations are not discriminated against by the strong in both North-South and South-South relations, but they argue that rule-making must proceed at a pace that is appropriate for the weakest members of the system, and the rules made in the WTO must be the right rules for development and poverty reduction. Current rules expose Northern governments to well-founded accusations of double standards on issues such as protection for domestic industries and support for domestic farmers, and are seen to provide insufficient

flexibility to enable Southern governments to pursue their development goals. The agenda being pushed by Northern countries is seen by many to militate against development and to be incompatible with the historical experience of the industrialized North.

Hegemonic Shifts In general terms, it is clear that significant changes have occurred in the thinking of policymakers since the mid-1980s. In part this has been a response to some of the more catastrophic results of gung-ho liberalization: the debacle of free market reforms in Russia, the Mexican crisis of 1994, and the Asian financial crisis of 1997 98 led to some serious soul-searching and admissions of mistakes, deflating the excessive self-confidence of the 1980s.

The growth of the anti-globalization movement fed off and accelerated this rethinking. Politicians recognized a need to respond to public disquiet, for example in the G8's decision to put debt on the agenda at its 1998 Birmingham summit, or when British prime minister Tony Blair used the 2005 Gleneagles G8 summit as a platform to gain further commitments for increased aid flows to Africa.

The emergence of a group of more economically powerful developing countries has also increased the political pressure for greater reform. For example, the creation of the G20 alliance of developing countries (including China, India, Brazil, and South Africa) at the WTO has shifted the balance of power within global trade relations and put greater pressure on the members of the Organisation for Economic Cooperation and Development to reform their agricultural sectors.

In recent years the movement has achieved some notable successes:

Jubilee 2000: This largely church-based coalition was credited by the British government with putting debt back on the international agenda. Initially started in the United Kingdom, Jubilee groups were set up in dozens of countries, North and South. Many, especially in the South, rapidly moved to campaign on wider globalization-related issues such as the impact of transnational corporations and structural adjustment programs.

Attac: This French-based network of intellectuals and activists has taken the lead in promoting the introduction of the Tobin tax (a small tax on currency transactions designed to curb speculative capital flows) and was influential in persuading the French government to support a study of the tax and oppose the Multilateral Agreement on Investment.

Corporate social responsibility: Public criticism and campaigning directed at corporate misconduct for example, pollution or abusive labor practices backed by increasing pressure from institutional investors, have prompted numerous initiatives to improve corporations' social and environmental performance. In the United States, student-led, grassroots, antisweatshop campaigns galvanized political life on campuses to a degree not seen since the Vietnam War.

These partial successes have strengthened the reformists within the movement and endangered its unity by heightening the points of difference between them and the rejectionists. The difficulties posed by partial victories were most clearly demonstrated in the Jubilee 2000 movement, when at the height of its policy successes at the Cologne G8 summit in 1999 the more radical "Jubilee South" wing, based in countries such as South Africa and Nicaragua, condemned the Northern Jubilee organizations for their reformist acceptance of the status quo.

Underlying the political debate has been a steady shift in public opinion, with messages on several fronts press exposés of poor working conditions, public protest, and the growing availability and prominence of "fair trade" products combining to make the public increasingly aware of the social impact of globalization.

It is easy to lose sight of how much has changed since the early 1980s in policy debates about globalization. There is now a much more nuanced understanding among decision-makers of the differences between liberalization of finance, direct investment, and trade; at the very least, most wings of the private sector pay lip service to notions of corporate social responsibility, and some of the most notorious excesses of free market zeal have been curbed.

The Future of Anti-globalization Understanding what may happen to this movement means exploring deeper questions about its political and social origins, the economic issues that it addresses, and the future of the international system to which it is a response. It also means looking at its component parts and their likely, and possibly differing, reactions to changing circumstances.

This brief overview suggests that much of this movement's coherence is contingent on external conditions. The strongest force in shaping its future development is therefore likely to be external, notably stemming from the pace and depth of change in the institutions of global governance and of the international system in which these institutions are based.

The success of the movement and in particular its reformist current in achieving change has been helped greatly by the multilateralism of the current international system. It has provided focus and coherence to otherwise disparate groups and allowed small gains in attempts to create a more balanced form of international governance.

Threats to multilateralism, however either from a turn toward unilateralism by the most powerful states or from the increased strain put on the long-term effectiveness of global governance by the exclusion of the world's weakest nations would strengthen the rejectionists and undermine the reformists, as well as the unity of the movement as a whole, and possibly the usefulness of the term itself.

See also fair trade; globalization; Group of Seven/Eight (G7/G8); Tobin tax; Washington consensus; World Economic Forum; World Trade Organization

FURTHER READING

Chang, Ha Joon. 2007. *Bad Samaritans: Rich Nations, Poor Policies, and the Threat to the Developing World.* London: Random House. A readable and critical examination of the policy prescriptions of the Washington consensus and an argument for the potential merits of active state involvement in successful development strategies that draws upon the experience of southeast Asia.

Desai, Meghnad, and Yahia Said. 2001. "The New Anti Capitalist Movement: Money and Global Civil Society."

In *Global Civil Society 2001*, edited by Helmut Anheier, Marlies Glasius, and Mary Kaldor. Oxford: Oxford University Press, 51 78. The Civil Society yearbook, based at the Centre for Civil Society at the London School of Economics, has been running since 2001 and pro vides a good overview of developments in the sector.

Fowler, Penny, and Kevin Watkins. 2003. *Rigged Rules and Double Standards: Trade, Globalisation, and the Fight against Poverty.* Oxford: Oxfam Academic. The bench mark publication that sets out the developmental case for fairer trade.

Green, Duncan, and Matthew Griffith. 2002. "Globaliza tion and Its Discontents." *International Affairs* 78 (1): 49 68. An overview of globalization and anti globali zation movements, from which this entry draws.

Held, David. 2004. *Global Covenant: The Social Democratic Alternative to the Washington Consensus.* Oxford: Polity Press. In many ways the counterpoint to Martin Wolf's case for the global economy, Held argues for a "social democractic alternative" with greater attention to pol icies to compensate those who lose out from globaliza tion, as well as the need for a stronger and more inclusive set of international institutions.

Klein, Naomi. 2001. *No Logo.* London: Flamingo. The best selling critique of global business trends, which blends cultural analysis, political manifesto, and jour nalistic exposé.

Mallaby, Sebastian. 2004. *The World's Banker: A Story of Failed States, Financial Crises, and the Wealth and Poverty of Nations.* New Haven, CT: Yale University Press. A highly readable portrait of James Wolfensohn's pres idency of the World Bank, which examines the struggles between the World Bank, its shareholders from the de veloped world, its clients in the developing world, and the NGOs that straddle both worlds.

Wolf, Martin. 2004. *Why Globalization Works: The Case for the Global Market Economy.* New Haven, CT: Yale University Press. A combative examination of the merits and faults of the different arguments from members of the anti globalization movement.

WEBSITES

Third World Network, http://www.twnside.org.sg
Focus on the Global South, http://www.focusweb.org
World Social Forum, http://www.worldsocialforum.org

Global Policy Forum, http://www.globalpolicy.org
Oxfam, http://www.oxfam.org.uk

DUNCAN GREEN AND MATT GRIFFITH

■ applied general equilibrium models

In their single-country form, applied general equi librium (AGE) models simulate the variety of markets that compose entire economies. In their regional or global forms, these models also simulate the market transactions among economies or groups of economies. In all these cases, AGE models ac count for the interactions in markets for goods, services, labor, physical capital, and other produc tive inputs. A large and growing body of evidence suggests that the indirect and economywide effects of policy changes captured in this framework are indeed important, and there is a well-established theoretical foundation for the approach. Economy-wide effects include upstream and downstream pro duction linkages through input-output relation ships; intersectoral competition for basic resources such as labor, physical capital, and land; realloca tion of rents from quantitative restrictions; distri butional effects across household types; and ex change rate changes. AGE models provide a means of explicitly accounting for these effects in a com prehensive and consistent manner.

The Notion of "Applied" What is meant by the notion of "applied" in AGE models? Francois and Reinert (1997) give the following four-part characterization:

Detailed policy orientation. Applied models use a broader set of policies than just ad valorem tariffs. The models involve an analytical commitment to the sectoral and institutional details of a policy, including the role of data nomenclatures and concordances among nomenclatures, as well as a commitment to the policymaking process itself and sensitivity to the kinds of results that are of interest to policymakers.

Nonlocal changes form distorted base equilibria. In most theoretical analyses of trade and other economic policies, the economy begins in a non-distorted state with no tariffs, quotas, or other taxes

present. To this initial nondistorted equilibrium an infinitesimal tariff or tax is introduced, and a new, counterfactual equilibrium is solved for using the linear approximation of differential calculus. In applied policy models, by contrast, the initial or base equilibrium reflects the relevant set of policy distortions. Since the policy changes introduced into the model are those actually under consideration rather than infinitesimal, the resultant changes are nonlocal. Consequently, functional forms prove to be crucial to model results, and what economists term "second-best effects" play a role in these outcomes.

Accurate and current data. Given the close link between modeling and policymaking, both the accuracy and the currency of data are important. Unfortunately, there can be a trade-off between accuracy and currency; the most recent data may be only estimates or not entirely survey based. Consequently, professional judgment is needed, and there needs to be a willingness to revise published results as new and better information becomes available.

Model structure determined by the data. The functional forms used to construct applied models must be chosen to allow for important regularities observed in the data. For example, the classical structures of theoretical trade models assume that imports and domestic competing goods are perfect substitutes and that there are no barriers to trade. In this case, the model describes interindustry trade only, cannot support a number of goods that exceed the number of factors of production, and cannot explain *bilateral* patterns of international trade. To avoid this in applied work, steps are taken to specify alternative model structures that allow for bilateral patterns of intraindustry trade and numbers of goods in excess of numbers of factors of production.

As an illustration of the complexities of applied policy modeling in general and the role of AGE models in particular, consider the case of multilateral Doha Round trade negotiations as described by Francois, van Meijl, and van Tongeren (2005):

Judging the economic impact of a WTO agreement is massively complex, even when it comes to issues as straightforward as tariff cutting. The eventual Doha Round agreement should lower thousands of individual tariffs in each WTO member country and there are about 150 members. The result would be important shifts of resources among sectors in most nations in the world, along with attendant changes in the prices of goods and productive factors. Due to international trade, the supply and demand factors in each nation affect resource allocations in all other nations. How can economists evaluate the impact of these choices? The most practical way of proceeding is to employ a large-scale computable general equilibrium model that allows simultaneous consideration of all the effects. (352)

This speaks to the utility of AGE models and explains why the methodology has evolved from an obscure exercise in applied econometrics to being a key tool for applied policy analysis. Indeed, many national governments use single-country models to assess policy changes, and both national governments and multilateral financial institutions use global models (e.g., the GTAP model described in Hertel 1997 or variants of the linkage model described in van der Mensbrugghe 2006) to assess a large variety of international policy issues.

Overview of Model Structure In AGE models, it is generally the case that an equation system is solved for prices that equate supply and demand in all markets simultaneously and satisfy the accounting identities governing economic behavior. Exceptions to this approach exist that allow for disequilibria in some markets (e.g., labor markets). Once the equation system has been specified, the equilibrium is calibrated to a base-year data set (often in the form of a social accounting matrix, or SAM), and the model reproduces the base-year economy in the absence of any policy changes. The calibration ensures that subsequent policy simulations move from an initial position that describes the economy (be it national, regional, or global) and its accounting identities as accurately as possible.

After calibration, the AGE model is used to simulate the effects of alternative policy changes

on the economy with reference to the base year (static) or a baseline scenario (dynamic). The model fully captures the flow of income from firms to labor, capital, and other inputs, from each of these inputs to households, and from households back to final demand of various types (household, government, investment, and rest of world or exports). Hence, it remains internally consistent even after introducing policy changes of various kinds.

The policy parameters of an AGE model include tariff rates, quotas, various domestic taxes and subsidies, and a number of behavioral elasticities. The behavioral elasticities play a central role in model results, and this represents an area of inquiry in which there is not always complete agreement with regard to appropriate values. Consequently, this is one aspect of AGE modeling that needs to be well documented and handled with care, including the analysis of the sensitivity of model outcomes to assumed elasticity values. Critical in this regard is what is known as the Armington (1969) elasticity, the elasticity of substitution between imports and domestic competing goods.

With a calibrated model in hand, an analyst can simulate the effects of proposed policy changes by comparing the base-year or baseline model solution with a counterfactual solution in which one or more of the policy parameters have been changed. A comparison of the base-year/baseline and counterfactual equilibria can reveal (depending on model specification) the effects of the policy changes on imports, exports, domestic production, employment, wages, aggregate economic welfare, disaggregated household welfare, and poverty and human development outcomes.

Standard Framework A standard framework of applied general equilibrium modes is that described by de Melo and Robinson (1989) and related to the previous contribution of Hazari, Sgro, and Suh (1980) on nontraded goods. In this framework, there are three varieties of goods: an export good (E), an import good (Z), and a nontraded, domestic good (D). Corresponding to these three goods, there are three different prices, P_E, P_Z, and P_D, and three dif-
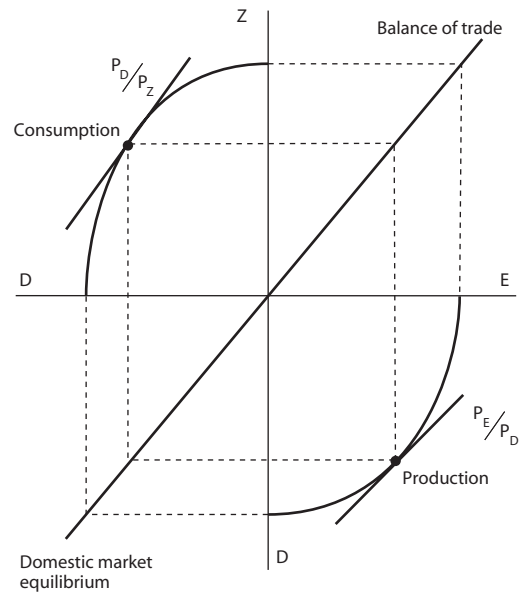
Figure 1
A graphical representation of a simple applied general equilibrium model

ferent quantities, E, Z, and D. The graphical representation of this framework is provided in the four quadrants of figure 1. In the southeast quadrant, the country produces two goods, E and D. The concave curve in this quadrant is the country's production possibilities frontier (PPF) and indicates that there are no economies of scale sufficient to outweigh the increasing opportunity costs of production. In practice, this part of AGE models is actually implemented in two stages, with a constant elasticity of substitution PPF and a constant elasticity of transformation allocation between exports and domestic supply, the latter following Powell and Gruen (1968). Under perfect competition, profit maximization, and full employment of resources, the country produces at a point on the PPF that is tangent to the relative price line P_E/P_D.

In the northeast quadrant of figure 1, we have a balance of trade line, which relates E and Z. This line tells us what level of imports can exist given the level of exports. As drawn, this balance of trade line is a 45-degree line, but changes in the terms of trade P_E/P_D will rotate the line either up or down. Capital inflows

or outflows will shift the line up or down, respectively, without changing its slope.

In the southwest quadrant of figure 1, the vertical axis plots the supply of the domestic good from the PPF, while the horizontal axis plots the demand for the domestic good. The 45-degree line in this quadrant is simply a domestic market equilibrium condition.

Finally, the northwest quadrant of figure 1 addresses the consumption of goods Z and D. The concave curve in this quadrant is a consumption possibilities frontier (CPF). The consumption possibilities are determined by the PPF, the balance of trade constraint, and domestic market equilibrium. The choice of the point on the CPF will reflect utility maximization in the form of tangency with an indifference curve of a representative household based on product differentiation by country of origin and is typically modeled using a constant elasticity of substitution frameworks as in Armington (1969).

In standard trade theory with incomplete specialization and no transportation costs, bilateral trade patterns remain indeterminate. However, in the framework of figure 1, intraindustry trade is fully captured. In multicountry versions of the model, actual patterns of bilateral, intraindustry trade will be effectively described in both base and counterfactual equilibria. Finally, nontradability in the framework of figure 1 depends on the price line P_E/P_D. The higher is this ratio, the more of domestic output will enter into trade. Therefore, nontradability in the AGE framework is not fixed by sector but depends on relative prices. Further, unlike Salter (1959), the two traded goods, E and Z, are not aggregated into a single good, and this is more satisfactory for empirical implementation.

Dynamics The policy changes considered in AGE models have the potential to change both overall income levels and the returns to factors of production. As noted by Francois, McDonald, and Nordström, "To the extent that investment hinges on income levels and expected returns, the medium- and long-term results of changes in trading conditions will include induced shifts in the capital stock" (1997,

365). Incorporating these considerations into AGE models shifts the analysis from *static* to *dynamic* considerations. Dynamic specifications of AGE models can vary in a number of respects, but the research on these specifications is now fairly well developed.

When allowed for, dynamic effects show up in model results through a number of paths, including the following: increases in incomes due to trade liberalization and other policy changes raising savings rates, declines in the prices of capital goods increasing real investment, and increasing trade openness (typically measured by sectoral export-to-output ratios) causing sectoral productivity increases. The first and second of these dynamic gains are often referred to as "medium-run growth effects" as identified by Baldwin (1992) and do not depend on any dynamic externalities as emphasized in New Growth Theory. Therefore, they are best viewed as reflecting "classical" or "old" growth theory. In this respect, the first and second paths for dynamic gains are quite firmly reflective of received economic theory.

The third type of dynamic effect is often referred to as the "procompetitive effects of exporting" and can reflect economies of scale in production and trade, learning by doing, standards achievement, and network effects. These potentialities are less standard than the medium-run growth effects, and it is not clear whether they are as automatic as suggested by some model specifications. In addition, there is little empirical evidence with regard to the behavioral elasticities used to model these effects. For these reasons, it is prudent to separate out simulations based on these procompetitive effects as indicative of *potential* rather than as actual, realized gains.

Other Extensions The standard AGE model considered here has been extended in a number of relevant directions. These include environmental and natural resource issues (including climate change), industry structure (including imperfect competition), migration, capital flows (including foreign direct investment), financial crises, commodity prices, foreign aid, natural disasters, poverty, and human development. Given this broad coverage of issues that are paramount in assessing the world economy, as well as the more standard applications

to trade policies, it is clear that AGE models will continue to be an important methodology to help us understand the role of national and multilateral policies in the world economy.

See also capital accumulation in open economies; gravity models; intraindustry trade; nontraded goods; partial equilibrium models; Swan diagram; terms of trade

FURTHER READING

Armington, Paul. 1969. "A Theory of Demand for Products Distinguished by Place of Production." *IMF Staff Papers* 16 (3): 159 76. The key contribution on product differentiation by country of origin used in most AGE modeling.

Baldwin, Richard E. 1992. "Measurable Dynamic Gains from Trade." *Journal of Political Economy* 100 (1): 162 74. A key contribution on the dynamic gains from trade.

de Melo, Jaime, and Sherman Robinson. 1989. "Product Differentiation and the Treatment of Product Differentiation in Computable General Equilibrium Models of Small Economies." *Journal of International Economics* 27 (1 2): 47 67. An introduction to the basic structure employed in many AGE models.

de Melo, Jaime, and David Tarr. 1992. *A General Equilibrium Analysis of U.S. Foreign Trade Policy*. Cambridge: Cambridge University Press. A thorough analysis of trade policy in a single country using the AGE framework.

Francois, Joseph F., Bradley J. McDonald, and Håkan Nordtröm. 1997. "Capital Accumulation in Applied Trade Models." In *Applied Methods for Trade Policy Analysis: A Handbook*, edited by Joseph F. Francois and Kenneth A. Reinert. Cambridge: Cambridge University Press, 364 82. An introduction to dynamic issues in AGE modeling.

Francois, Joseph F., Hans van Meijl, and Frank van Tongeren. 2005. "Trade Liberalization in the Doha Development Round." *Economic Policy* 20 (42): 349 91. An application of AGE modeling to multilateral trade negotiations.

Francois, J. F., and K. A. Reinert. 1997. "Applied Methods for Trade Policy Analysis: An Overview." In *Applied Methods for Trade Policy Analysis: A Handbook*, edited by Joseph F. Francois and Kenneth A. Reinert. Cam

bridge: Cambridge University Press, 3 24. A well known introduction to the field of applied trade policy analysis.

Harrison, Glenn W., Thomas F. Rutherford, and David G. Tarr. 2003. "Trade Liberalization, Poverty, and Efficient Equity." *Journal of Development Economics* 71 (1): 97 128. An example of using household survey data in an AGE model to assess the distributional impacts of trade liberalization.

Hazari, Bharat R., Pasquale M. Sgro, and Dong C. Suh. 1980. "A Simple Geometric Treatment of Nontraded Goods in the Pure Theory of International Trade." *Keio Economic Studies* 17 (1): 91 104. A trade theoretic treatment of nontraded goods that corresponds to a framework adopted in many AGE models.

Hertel, Thomas W., ed. 1997. *Global Trade Analysis: Modeling and Applications*. Cambridge: Cambridge University Press. A detailed description of a widely used, global AGE model known as GTAP.

Markusen, James R., Thomas F. Rutherford, and Linda Hunter. 1995. "Trade Liberalization in a Multinational Dominated Industry." *Journal of International Economics* 38 (1/2): 95 117. An application of AGE modeling to foreign direct investment.

Powell, A., and F. Gruen. 1968. "The Constant Elasticity of Transformation Production Frontier and Linear Supply System." *International Economic Review* 9 (3): 315 28. The source for the functional form used in modeling the allocation of goods between exports and domestic supply in AGE models.

Reinert, Kenneth A., and David W. Roland Holst. 1997. "Social Accounting Matrices." In *Applied Methods for Trade Policy Analysis: A Handbook*, edited by Joseph F. Francois and Kenneth A. Reinert. Cambridge University Press, 94 121. An introduction to social accounting matrices and their use in AGE models.

Salter, Wilfred E.G. 1959. "Internal and External Balance: The Role of Price and Expenditure Effects." *Economic Record* 35 (171): 226 38. A classic treatment of traded versus nontraded goods that has proved important in AGE modeling.

Shoven, John B., and John Whalley. 1984. "Applied General Equilibrium Models of Taxation and International Trade: An Introduction and Survey." *Journal of Eco*

nomic Literature 22 (3): 1007 51. An early and noted review of AGE models in both trade policy and public finance.

van der Mensbrugghe, Dominique. 2006. *LINKAGE Technical Reference Document, Version 6.0.* Washington, DC: World Bank. A detailed description of a widely used, global AGE model that emphasizes development issues.

KENNETH A. REINERT

■ appropriate technology and foreign direct investment

Appropriate technology generally refers to technology that is suitable for local conditions. Although the term has sometimes been used very broadly to refer to technologies that support goals such as the preservation of nature, local culture, or life in small villages, economists typically speak of technological "appropriateness" in the context of technologies adapted for local conditions such as factor abundance for example, capital-intensive technologies in capital-abundant countries and labor-intensive technologies in labor-abundant countries.

The discussion of appropriate technology in the context of foreign direct investment (FDI) arose from concerns that technologies developed under the economic conditions of developed countries could have harmful effects in developing countries in particular, that multinational enterprises (MNEs) would not create enough jobs. This possibility is particularly likely if innovations are "induced" by cost-saving possibilities in the markets where they are first used (Kennedy 1964). Because developed countries are capital-abundant, the technologies developed there tend to be capital-intensive, labor-saving technologies, and thus tend to generate less employment in a developing country than do technologies tailored to the conditions of abundant, cheap labor present there. Similar arguments apply to other features of technology: for example, technologies requiring a steady supply of electric power may be inappropriate for countries that experience fre-

quent power outages. Thus there has been a general concern whether MNEs bring in the right type of technologies whether the technologies that are in the best interest of MNEs to transfer are those that are in the best interest of the local economy.

Adaptation and Technology Transfer In simple models of the behavior of MNEs, the technology developed in the home market is often assumed to be reproduced identically in other countries, which is how the firm exploits its advantage of technological ownership. In fact, reproducing technology abroad is not easy; technology often must be adapted to local conditions for it to work. If the existing technology is not appropriate for local conditions in the host country, the MNE may have an incentive to adapt the technology.

Actual technology transfer is not simply a matter of shipping blueprints to another country, but involves a costly process of experimentation and trial and error (Teece 1977), during which the specifications of the technology adapt somewhat to local conditions. Some technologies may be more easily adaptable than others, such as ones where different techniques using different mixes of capital and labor are feasible. In some cases the costs of adaptation may be large enough that adaptation is not worthwhile to the MNE. The practical questions thus become a matter of how adaptable MNEs' technologies are to local conditions, how much adapting MNEs in fact decide to do, and whether local firms would in fact do a better job of adaptation (Lall 1978).

The degree to which MNEs adapt their technologies to local conditions has been widely studied. Since adapting technology is costly, and since the competitive advantage of MNEs lies in large part in their ability to replicate technologies they own in different countries, there can be a strong incentive to minimize adaptations to local conditions. For example, the basic type of machinery is likely to stay the same from country to country. Thus technologies initially created for local conditions may well be better suited for the local conditions than technologies created elsewhere and then adapted. However, the situations where technologies are ill adapted are

apt to be those where adaptation is difficult, and hence local firms may be unable to do better than MNEs.

Empirical results on the observed degree of technological adaptation often defy easy generalization. Large statistical studies often include firms in different industries, and therefore the differences in capital intensity may be due to the industry rather than to individual firm choice. The effect of foreign firms' technological choices on domestic firms in developing countries depends in part on whether the local firms are suppliers, are competitors, or have some other relationship to foreign firms. It has been reported that MNEs impose particular technologies on their suppliers in developing countries, at least partially in an attempt to maintain quality control. There are many possibilities for developing country firms to learn and imitate the technology of MNEs, through reverse engineering, social networking among employees, or employees who leave to become entrepreneurs or are hired away by competitors. Sometimes, domestic imitators have been more successful adaptors than MNEs and have been rewarded by rising market shares. At the same time, one can observe cases in which managers in developing-country subsidiaries of MNEs from developed countries adapt technologies in the labor-intensive direction (Pack 1976; Lecraw 1977).

Multinationals from Developing Countries In the 1960s and 1970s, an increasing number of firms based in countries such as India, Hong Kong, Korea, Brazil, and Argentina became direct investors. These "third-world multinationals" tended to place the bulk of their investments in other developing countries. Since economists were accustomed to think of MNEs transferring capital and technology from the North to the South, new explanations seemed necessary for the phenomenon of South-South FDI. A widely offered idea was that developing-country firms had acquired "appropriate" technologies in their home markets, which gave their affiliates in other countries a competitive advantage.

The types of intangible firm advantages that could arise from operating in a developing-country environment are widespread. These could include managerial skills in dealing with severe bottlenecks in acquisition of materials and power, local labor conditions, and developing-country bureaucracies, and superior information about local product and factor markets in certain trading partners. Also, in some countries such as Korea and India, the largest domestic firms were organized as multiproduct conglomerates, and the managerial skills needed for such organization readily adapt to the multinational form.

The motivations for South-South FDI need not be limited to intangible appropriate-technology advantages. Informational advantages gained from a past history of exporting or migration suggest a transactions cost explanation rather than a technological one. Even in developing countries, firm-level efforts to promote efficiency might encourage capital intensity rather than labor intensity (Ferrantino 1992). By the early 21st century, some investors from developing countries were seeking to acquire developed-country technologies through mergers and acquisitions (e.g., Chinese investors in the United States), showing that home-country production experience need not be the only determinant of technological development in developing-country MNEs.

Appropriate Technology, Efficiency, and Spillovers From a policy standpoint, part of the idea of appropriate technology is that the technology chosen by markets might not be optimal for either economic growth or employment generation. This imperfection suggests a role for government either in technology choice or at least in undoing distortions such as those that might make labor artificially expensive. In the 1980s the skepticism about the benefits of FDI for developing countries began to wane, and the promarket ideas collectively called the Washington consensus tended to see FDI as enhancing productivity and growth rather than being an original source of distortions. Under this view, the most important feature of technologies bundled with FDI is their potential for enhancement of overall productivity, rather than their bias toward either capital or labor.

Nonetheless, the question of the conditions under which the activities of developed-country MNEs influence technological conditions in developing

countries, and whether and under what conditions adaptation takes place, remains open. Much of the current research on technological spillovers from FDI is informed by the issues raised in the analysis of appropriate technology.

See also foreign direct investment and innovation, imitation; foreign direct investment and international technology transfer; location theory; technology spillovers

FURTHER READING

Ferrantino, Michael J. 1992. "Technology Expenditures, Factor Intensity, and Efficiency in Indian Manufactur ing." *Review of Economics and Statistics* 74 (4): 689 700. Firm level study examining whether the technologi cal efforts of Indian firms, prominent among the early third world multinationals, led to "appropriate" out comes.

Kennedy, Charles. 1964. "Induced Bias in Innovation and the Theory of Distribution." *Economic Journal* 72 (295): 541 47. Early treatment of the response of in novation to relative resource prices.

Lall, Sanjaya. 1978. "Transnationals, Domestic Enterprises, and Industrial Structure in Host LDCs: A Survey." *Oxford Economic Papers* NS 30 (2): 217 48. A treatment of technological appropriateness and adaptation in the context of available evidence on linkages between transnationals and developing countries.

———, ed. 1983. *The New Multinationals: The Spread of Third World Enterprises.* Chichester: John Wiley and Sons. Case studies on multinationals from India, Hong Kong, Argentina, and Brazil.

Lecraw, Donald J. 1977. "Direct Investment by Firms from Less Developed Countries." *Oxford Economic Papers* NS 29 (3): 442 57. Demonstrates, for Thailand, that in vestors from developing countries have different moti vations from developed country investors and use more labor intensive technology.

Pack, Howard. 1976. "The Substitution of Labor for Cap ital in Kenyan Manufacturing." *Economic Journal* 86 (341): 45 58. Firm level study revealing that both do mestic and foreign firms were capable of improving productivity without capital deepening.

Teece, David J. 1977. "Technology Transfer by Multi national Firms: The Resource Cost of Transferring Technological Know how." *Economic Journal* 87 (346): 242 61. Evidence on the level and determinants of technology transfer costs within MNEs.

MICHAEL FERRANTINO

■ arbitrage

See interest parity conditions

■ Asia Pacific Economic Cooperation (APEC)

The Asia Pacific Economic Cooperation (APEC) process was the first structured forum for intergovernmental cooperation among the economies of East Asia and their main trading partners in North America, Australasia, and Russia. It was launched in Canberra, Australia, in 1989 to promote closer communications to help seize the opportunities for trade and investment among a diverse group of economies with different resources and comparative advantages, as well as to anticipate the inevitable tensions when comparative advantage changed as economies developed. APEC's central objective is to sustain the growth and development of the region by promoting mutually beneficial economic integration and by encouraging international flows of goods, services, capital, and technology. APEC has adopted the principle of open regionalism, seeking to reduce impediments to international economic transactions among participants without diverting economic activity away from other economies.

APEC was built on foundations laid during the preceding decades. Since the 1960s, the Association of Southeast Asian Nations (ASEAN) has demonstrated that a voluntary association of diverse nations with diverse economies can be valuable and effective. Since those years policy-oriented analysis by researchers and business people with a broad international outlook noted the growing, market-driven interdependence of Asia Pacific economies. This included the work of the Pacific Forum for Trade and Development, a group of policy-oriented researchers that has met annually since 1967 to assess the changing environment for economic development

in the Pacific and its policy implications; the Pacific Basin Economic Council, a group of senior business people with a broad international outlook that also began to meet in the late 1960s; and the Pacific Economic Cooperation Council, established in 1980. APEC has also set up its own private-sector advisory group, the APEC Business Advisory Council.

In a January 1989 speech in Seoul, Australian prime minister Bob Hawke advocated the creation of a new intergovernmental vehicle of regional cooperation. The first ministerial-level meeting brought together representatives from twelve Asia Pacific economies. The initial participants were the then six members of ASEAN (Brunei, Indonesia, Malaysia, the Philippines, Singapore, and Thailand) together with Australia, Canada, Japan, the Republic of Korea, New Zealand, and the United States. Subsequently, APEC developed a comprehensive work program to promote economic integration and agreed on its main objectives and principles set out in the 1991 Seoul APEC Declaration.

APEC further distinguished itself from the legally binding and preferential arrangements adopted by the European Union (EU) and others by agreeing that Asia Pacific cooperation should be voluntary. APEC would not be a negotiating forum: instead it would seek to identify opportunities to promote mutual economic benefit through consultation, building consensus around a widening range of shared interests. APEC's nonformal structure made it possible to include all three Chinese economies. Taiwan participates as the economy of Chinese Taipei, and since 1997 Hong Kong has participated as Hong Kong, China. Membership has since expanded to 21: Chile, Mexico, Papua New Guinea, Peru, Russia, and Vietnam joined between 1993 and 1997. A subsequent moratorium on expanded membership was set to end in 2007.

Free and Open Trade and Investment APEC leaders met for the first time in 1993. At their 1994 meeting, in Bogor, Indonesia, they were able to make a political commitment to eliminating barriers to trade and investment by no later than 2010 for de-

veloped economies and 2020 for developing economies. A midterm stocktaking, conducted in 2005, showed considerable progress toward this goal. Average tariffs were considerably lower than in 1989 and border barriers to trade in most goods and many services were already set at zero or negligible levels. People and capital were moving much more freely around the region.

By harmonizing customs procedures, increasing the scope of mutual recognition of standards, adopting agreed principles for more transparent and competitive government procurement, and reducing impediments to international business travel, APEC made substantial progress in reducing other costs and risks of international commerce. This progress has led to the emergence of the Asia Pacific region as the engine of global economic growth, outpacing the rest of the world in terms of opening itself to international trade and investment and increasing its share of global output, trade, and investment (APEC 2005; Elek 2005).

Strengths and Weaknesses APEC's experience has also revealed some of the weaknesses as well as the strengths of voluntary cooperation among widely diverse economies. Coordinated unilateral actions have helped to bring down many impediments to trade and investment, but voluntary cooperation has not proved adequate to liberalize "sensitive sectors," such as agriculture, textiles, and clothing. APEC's limited ability to respond to the serious financial crises in East Asia in the late 1990s also demonstrated the value of attending to all of the foundations of sustained economic growth, not just the reduction of obstacles to international trade and investment.

By 2007, it was evident that one dimension of the vision of free and open trade and investment, the elimination of all traditional border barriers to trade, would not be achieved according to the 2010/2020 timetable. At the same time, there was growing realization that the removal of tariffs or quantitative restrictions on trade in goods, or even in services, would not be enough to promote deep economic integration. The experience of the EU had shown the value of a comprehensive program to reduce reg-

ulatory impediments to international movement, not only of products, but also of factors of production.

Facilitating economic integration by closer coordination or harmonization of policies on matters such as customs procedures or standards, adoption of compatible policies to encourage e-commerce, and reduction of the cost of international transportation have become relatively more important means of promoting economic integration. In these areas, the effective constraint on cooperation is not the short-term political cost of overcoming narrow vested interests but the capacity to implement more efficient policies. Facilitating trade by dealing with logistic or regulatory obstacles to trade is largely a matter of enhancing human, institutional, and infrastructure capacity.

The Busan roadmap for promoting progress toward free and open trade and investment, based on the 2005 midterm evaluation, takes account of these issues. Accordingly, the focus of attention is shifting from the remaining traditional border barriers to trade toward the sharing of information, experience, expertise, and technology to help all Asia Pacific economies build the capacity for designing and implementing better policies, which strengthen domestic as well as just international markets. At the same time, APEC economies would support traditional trade liberalization through the WTO. This could bring about a practical division of effort between APEC and the WTO, based on the comparative advantages of institutions for voluntary, rather than negotiated, cooperation.

Especially since 2000, APEC has needed to define its role alongside many other bilateral and subregional forms of economic cooperation. In response to the inability to make further significant progress on sensitive issues, such as agriculture, in either the WTO or APEC, combined with the need to address many new issues that influence international economic transactions, there has been a proliferation of mostly bilateral preferential trading arrangements. Yet bilateral preferential trading arrangements have also been unable to achieve significant liberalization of trade in sensitive products, while complicating all

international trade due to complex, and often overtly discriminatory, rules of origin. It remains to be seen whether that problem can be overcome by preferential trading arrangements among larger groups of economies.

Concurrently, reflecting the growing share of international production and trade in East Asia, several East Asia centered forums have emerged. As in APEC, members of the wider networks of cooperation in East Asia will resist ceding powers to any supranational authority. This suggests that they might best focus on issues of perceived shared interests, rather than the negotiation and imposition of binding constraints on one another.

In summary, APEC members have moved toward their original goals but their institution has been redefining its own goals and operation. At the same time, other institutional developments in the region now create the prospect of competition and overlap with APEC (Soesastro and Findlay 2005). Progress on economic integration may prove easier in an East Asian, rather than in a wider trans-Pacific, forum. On the other hand, East Asia is itself a very diverse grouping and will face constraints quite similar to those encountered by APEC while the United States will seek to sustain trans-Pacific cooperation. One outcome may be "variable geometry," with APEC-wide consultations encouraging different groups of economies to pioneer cooperation on issues of shared interest, not necessarily in the institutional context of the APEC process itself.

See also Association of Southeast Asian Nations (ASEAN); free trade area; regionalism

FURTHER READING

APEC. 2005. "A Mid term Stocktake of Progress to ward the Bogor Goals Busan Roadmap to Bogor Goals." (November) Busan, Korea: APEC. Down loadable from http://www.apec.org/apec/enewsletter/ jan vol7/publication.primarycontentparagraph.0009. A midterm report by APEC officials, based on research by a panel of economists from the region.

Elek, Andrew. 2005. "APEC after Busan: New Direction." APEC Study Series 05 01 (November) Seoul: Korean

Institute for International Economic Policy. Down loadable from http://www.kiep.go.kr/eng/std data view.asp?num＝131907&sCate＝013002&sSubcate＝&lTp＝r&nowPage＝3&listCnt＝15#. This paper sets out the reasons for adopting APEC's 2005 Busan road map and how that agenda could be realized.

Soesastro, Hadi, and Christopher Findlay, eds. 2005. *Res haping the Asia Pacific Economic Order*. London: Rout ledge. A collection of papers including a review of APEC's contribution and analyses of the com plementarities and conflicts among institutions in the region; includes a note on the contribution of Peter Drysdale to the development of the APEC concept.

ANDREW ELEK AND CHRISTOPHER FINDLAY

■ **assignment problem**

The assignment problem concerns the allocation of policy instruments to policy targets in order to im prove policy effectiveness. Policy instruments are the variables or procedures that policy authorities di rectly control. Policymakers' use of these instruments to achieve objectives (i.e., policy targets) directly af fects the welfare of their constituents. For example, government spending is an instrument that can be used to reduce unemployment. Assignment is im portant because in actual policymaking, competing government agencies can be given different instru ments. The question then becomes which is the appropriate allocation?

Trevor Swan (1960) noted the problem in the context of an open economy where the two objectives were full employment with price stability and a rea sonable current account deficit (CAD) of the balance of payments. The first is the objective of internal balance and the second that of external balance, which is an essential part of participating in the world economy. External imbalances in one country can lead to imbalances for others with which it transacts. One country's delays and inefficiencies in achieving balance impinge on others, and when the country is large, can have repercussions for the global financial system.

The Tinbergen Rule In 1952, Jan Tinbergen demonstrated that to achieve policy objectives, gov ernments must have policy instruments equal in number to the objectives. According to what is known as the Tinbergen rule, a government cannot achieve two objectives with just one instrument. If a country has unemployment and a CAD, for exam ple, and the only instrument available is government spending, it can fix only one problem (Meade 1951). If the government spends more to stimulate the economy, it will lower unemployment but worsen the disequilibrium in the balance of payments as overall output and imports rise. If it reduces spend ing, the balance of trade improves, but unemploy ment rises. One objective is attained at the expense of the other. If the government uses two instruments government spending and exchange rate policy to achieve its objectives, the question is one of appro priate assignment: which instrument should it assign to control demand, output, and employment, and which to achieve trade balance?

Adjustment in the Swan Diagram Swan (1960) showed that an incorrect assignment can move the economy away from full equilibrium. Figure 1 shows a simple Swan diagram where the internal balance, or domestic market equilibrium schedule, IB_f is downward sloping in the space of real exchange rate
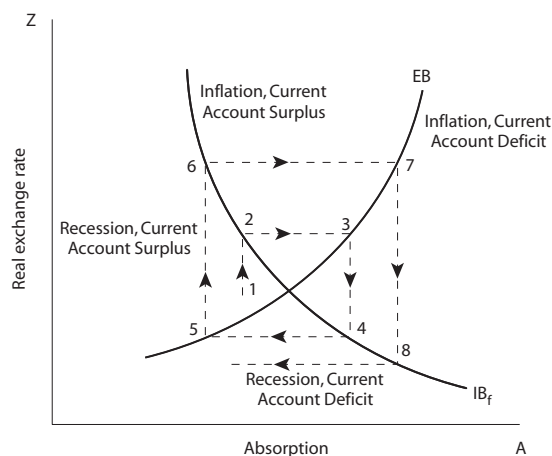
Figure 1
Assignment in the Swan diagram

Z and absorption, or domestic demand, A. Z is the ratio of external to domestic costs of production. Currency appreciation reduces demand for domestic goods so that government spending, a component of absorption, has to rise to maintain output unchanged. The external balance, or balance of payments equilibrium schedule, EB, is upward sloping. The current account will improve with real depreciation as imports fall and exports rise, so government spending must be raised to restore imports and prevent the balance of payments going into surplus. The four quadrants denote different types of disequilibria. Only at the point of intersection of the two schedules are the two instruments such that both the targets are achieved. Output is demand determined below the full employment IB_f schedule and otherwise limited by the factor of production labor.

Consider an economy at point 1 with a current account surplus and a recession. Policymakers assign the real exchange rate to achieve internal balance and government spending to achieve external balance, and sequentially adjust the two instruments. Depreciation at point 1 will lead to internal equilibrium at point 2 as exports raise demand. A rise in government spending raising imports to match will bring the economy to external equilibrium at point 3, and so on through points 3, 4, 5. . . . But instead of converging to the full equilibrium point, the economy diverges away from it. The opposite assignment leads to equilibrium. Reversing the arrows before point 8 shows the convergence to full equilibrium where the starting point is a recession and a CAD. Government spending then is raised to reach internal balance at point 8, and then the exchange rate depreciated to reach external balance at point 7, the process converging finally to the full equilibrium. Thus this is the correct assignment of instruments.

Note that the result depends on the relative steepness of the two schedules. Convergence to a stable equilibrium occurs because the internal balance schedule is steeper than the external balance. Depreciation has more of an effect on the current account than it does on aggregate demand since changes in imports and exports act in the same direction on the CAD, but imports are a leakage from demand for domestic goods, and only the rise in exports raises output. Since a larger change in absorption is required to compensate for the relatively greater impact of depreciation on external balance, the EB schedule is flatter than the IB_f. That is, the exchange rate has a relatively greater impact on the external balance than it does on the internal balance. And government spending has a greater impact on internal balance than it does on external balance. It follows that exchange rate policy should be assigned to external balance, and government spending to internal balance.

Effective Market Classification Each policy instrument should be assigned to the target variable on which its relative effect is higher. Robert Mundell first formally stated this in 1962 as the principle of effective market classification. Consider a case where the two policy instruments are the interest rate and the budget deficit, with a fixed exchange rate. If the rate of interest has a larger effect on external balance, Mundell showed that the interest rate must be assigned to external balance and the budget deficit to internal balance in sequential adjustment; otherwise the economy will diverge from full equilibrium. The interest rate affects both the current and the capital account. Suppose government expenditure and interest rates rise, moving away from the unique equilibrium, but with internal balance satisfied. The rise in government expenditure tends to raise income, but the rise in interest rate reduces it, so that income remains at the full employment level. Since output is unchanged, the trade balance is unchanged as well, assuming exports are constant and imports depend only on income. But the rise in interest rates improves the capital account of the balance of payments, so that the balance of payments is now in surplus. Therefore the internal balance schedule must be steeper than the external balance schedule.

Starting from a position with recession and a CAD, if the government reduces spending to reduce the CAD, it will improve the CAD but worsen the recession. If it then decreases the interest rate to stimulate output, it will worsen the capital account as

capital flows out, resulting in cumulative movements away from equilibrium. Adjustment converges to the full equilibrium if the opposite assignment is made. That is, the interest rate is directed to achieve external equilibrium and government spending to achieve internal equilibrium each policy instrument is assigned to the target on which it has the greatest effect, or which it has a comparative advantage in achieving. In this simple model with fixed exchange rates, the implication is that monetary policy or the central bank should target external balance, while fiscal tax-expenditure decisions target full employment.

In practice, sequential policy adjustment is difficult since the economic system is simultaneous. In both models, since the full equilibrium is stable, it is attained with simultaneous adjustment of both instruments. Markets themselves also generate adjustment in response to disequilibrium. But optimal assignment is a useful feature to keep in mind when making a complex policy decision, especially if the two instruments are the responsibility of different institutions. The correct allocation of separate instruments to separate authorities may allow more efficient adjustment. In addition, since all models are a simplification of a complex reality, as international and domestic institutions change, the relevant model also changes or must be suitably adapted.

Allowing for Shocks In the simple deterministic models considered so far, there were no random shocks. Poole (1970) has an interesting variant of the assignment problem regarding the choice of the appropriate instrument for monetary policy. He showed that if aggregate demand was subject to more random shocks compared to the demand for money, money supply should be the operating instrument. Output fluctuations would then be lower, since procyclical interest rate movements would moderate the shocks. If the intensity of shocks to money demand was greater, interest rates should be the instrument, since automatic adjustment of money supply would smooth interest rates, reducing output fluctuations. The opposite assignment would magnify the impact of fluctuations.

See also balance of payments; comparative advantage; discipline; exchange rate regimes; expenditure changing and expenditure switching; money supply; Mundell-Fleming model; structural adjustment; Swan diagram; twin deficits

FURTHER READING

Agenor, Pierre Richard, and Peter J. Monteil. 1999. *Development Macroeconomics*. 2d ed. Princeton, NJ: Princeton University Press. Detailed case studies of developing country experiences, which illustrate problems with the assignment of macroeconomic policy instruments.

Corbo, Vittorio, and Stanley Fischer. 1995. "Structural Adjustment, Stabilisation, and Policy Reform: Domestic and International Finance." In *Handbook of Development Economics*, edited by Jere Behreman and T. N. Srinivasan. Vol. 3. Amsterdam: North Holland, chapter 44. An application of the Swan framework to modern stabilization and structural adjustment programs, which can be understood as assignment problems.

Meade, James E. 1951. *The Theory of International Economic Policy*. Vol. 1. *The Balance of Payments*. London: Oxford University Press. A classic of international economic policy.

Mundell, Robert A. 1962. "The Appropriate Use of Monetary and Fiscal Policy for Internal and External Stability." *IMF Staff Papers* 12 (9): 70 79. Examination of the dynamics of inappropriate assignment.

Poole, William. 1970. "Optimal Choice of Monetary Policy Instruments in a Simple Stochastic Model." *Quarterly Journal of Economics* 84 (2): 197 216. The effect of shocks on choice of policy instruments.

Swan, Trevor W. 1960. "Economic Control in a Dependent Economy." *Economic Record* 36 (March): 51 66. The seminal discussion of assignment.

Tinbergen, Jan, 1952. *On the Theory of Economic Policy*. Amsterdam: North Holland. The original development of the Tinbergen rule.

van Marrewijk, Charles. 2007. *International Economics: Theory, Application, and Policy*. London: Oxford University Press. Simple clear textbook treatment of the assignment problem.

ASHIMA GOYAL

■ Association of Southeast Asian Nations (ASEAN)

The Association of Southeast Asian Nations (ASEAN) is an intergovernmental association that was formed to enhance cooperation among countries in the Southeast Asian region. Since its formation in August 1967, this association has grown in size to include 10 countries: Brunei, Cambodia, Indonesia, the Lao People's Democratic Republic (PDR), Malaysia, Myanmar, the Philippines, Singapore, Thailand, and Vietnam.

The formation of ASEAN was officially opened at the signing of the Bangkok Declaration on August 8, 1967, by five countries: Indonesia, Malaysia, the Philippines, Singapore, and Thailand. Brunei in 1984, Vietnam in 1995, the Lao PDR and Myanmar in 1997, and Cambodia in 1999 joined the grouping. Political and economic considerations have influenced these countries to form a regional cooperation. The original five signatory countries saw the need to foster their economic development and promote regional security in the face of a growing communist threat in Southeast Asia, precipitated by the fall of Indochina to communism and the declared intention of the West to withdraw its military forces from the region. Their common objectives could be best achieved through mutual cooperation in the economic, social, and cultural areas.

Political Cooperation During its early stages of development, ASEAN was mainly concerned with political cooperation to promote political stability and harmony in the region. Since 1976, however, the economic objective has grown in importance as economic development has become its main priority in view of, among other things, the increasing need to maintain economic competitiveness in the midst of growing international competition and sustain member countries' pace of economic development in light of rising expectations.

Economic Cooperation There have been two types of economic cooperation in ASEAN: market sharing and resource pooling. These types of economic cooperation have developed since the Bali Summit of 1976. At this summit, the ASEAN leaders signed a Declaration of ASEAN Concord that laid out a program of action for regional cooperation in political, economic, social, cultural, and security matters. But ASEAN economic cooperation of a market-sharing type has generally been unsuccessful mainly due to the significant differences in members' economic development levels and different economic priorities.

Economic Cooperation in Trade The most important form of market-sharing cooperation envisaged in the Bali Summit was the establishment of preferential trading arrangements to promote intra-ASEAN trade. The agreement on ASEAN Preferential Trading Arrangement, signed in February 1977, was the first attempt by ASEAN member countries to promote a higher level of intra-ASEAN trade. Five measures have been identified to achieve the objective of greater intra-ASEAN trade: exchange of tariff preferences, purchase of finance support at preferential rates for selected products of ASEAN domestic origin, long-term (three to five years) quantity contracts for basic commodities such as fuels and agricultural products, preference in procurement by government agencies, and liberalization of nontariff barriers on a preferential basis.

Economic Cooperation in Industry ASEAN countries have resolved that industrialization and economic development of their members can be largely facilitated through industrial cooperation. There is also a common realization that a significant increase in intra-ASEAN trade can occur only if the supply side of the market is also increased. Thus since 1976 five major schemes of ASEAN industrial cooperation (ASEAN Industrial Projects Scheme, ASEAN Industrial Complementation Scheme, Brand-to-Brand Complementation Scheme, ASEAN Industrial Joint Venture Scheme, and ASEAN Industrial Cooperation Scheme) have been established to create new industrial capacity jointly owned by member countries serving the regional market. Unlike trade liberalization, which raises apprehensions of uncontrolled trade flows and market disruption, industrial cooperation is linked to the creation of specific new production facilities, giving rise to much more predictable trade flows. Except for the ASEAN

Industrial Cooperation Scheme, the success of these initiatives was quite limited.

Economic Cooperation in Agriculture Given the importance of the agricultural sector in ASEAN, cooperation in this area has been one of the priority issues. All the cooperative undertakings in agriculture have so far been technical and developmental in nature, however. The Committee on Food, Agriculture, and Forestry (COFAF) is responsible for the identification and implementation of cooperative undertakings in the agricultural sector. Formed in March 1977, it (a) coordinates, reviews, and prepares studies on the prospects of the agricultural sector; (b) develops efficient methods for the exchange of information on agriculture; (c) identifies areas for cooperation; (d) maintains close ties with related committees in ASEAN and with other related organizations inside and outside the region; and (e) reports its progress to the ASEAN Economic Ministers.

Between 1978 and 1988, COFAF had identified 46 cooperative projects—7 in food, 10 in agriculture, 15 in forestry, 8 in livestock, and 6 in fisheries. By 1985, 13 of them were ongoing, 4 completed, and 29 not yet implemented. The largest number of projects was located in the Philippines, followed by Malaysia, Thailand, and Indonesia in descending order. Most of the funding for these projects came from sources outside the region and only a few projects were pursued on a long-term basis. One of them was the ASEAN Food Security Project, which established in 1979 an ASEAN Emergency Rice Reserve to be contributed to by each of the member countries. The objective was to create a stockpile of 50,000 metric tons of rice to meet shortfalls in domestic supply. This program is periodically reviewed for the purpose of creating a more dynamic food security arrangement to enhance intra-ASEAN trade and promote food production under the principle of comparative advantage. Others included the Food Handling and the Seed Technology projects. ASEAN cooperation in agriculture has also been fostered in the association's dealings with developed countries on matters of market access for member countries' primary exports, better terms of trade, stabilization

of export earnings, and other nonsensitive areas of cooperation.

ASEAN Free Trade Area The 1992 ASEAN Free Trade Agreement (AFTA) was a watershed in the history of ASEAN economic cooperation, as it represents a significant step in the economic policy orientation of the ASEAN countries. The main objective of AFTA is to increase the international competitiveness of ASEAN industries and the ASEAN region as an investment location. Specifically, the objectives are to increase intra-ASEAN trade by abolishing intraregional trade barriers while allowing member countries to keep their respective trade policies toward the rest of the world, attract local and foreign investors to invest in the region, and make their manufacturing sector more efficient and internationally competitive within a liberalizing global market. An integrated regional market is expected to produce economic benefits from greater consumer welfare, exploitation of economies of scale, competition-induced efficiency, industrial rationalization, interindustry linkages, and intraindustry trade.

To realize these benefits, AFTA seeks to reduce tariffs on all commodities traded within the member countries to no more than 5 percent ad valorem and remove all other trade restrictions by the year 2002 under the Common Effective Preferential tariff, the main instrument of AFTA. The agreement also lays down the rules for fair competition and identifies a number of measures to enhance economic cooperation such as harmonization of standards, macroeconomic consultations, improved reciprocal recognition of product testing and certification, coordination of foreign investment policies to enhance more investment flows, joint investment promotion strategies, and cooperation in transportation systems. Further, it contains measures of contingent protection and allows the reintroduction of trade barriers in case of balance of payments difficulties.

ASEAN Economic Community The Asian financial crisis of 1997 and 1998, the terrorist attacks of September 11, 2001, the Iraq war, and the SARS outbreak all dampened the outlook for greater trade liberalization and economic integration as envi-

sioned under AFTA. Despite these unfavorable external events, the ASEAN countries have remained politically committed to the vision of transforming ASEAN into an economically integrated grouping within the framework of an ASEAN Economic Community (AEC). This AEC concept, which was first agreed on at the November 2002 ASEAN summit held in Phnom Penh, Cambodia, is based on the ASEAN Vision 2020, which foresees a more economically integrated ASEAN. The ASEAN Vision 2020 envisioned "a stable, prosperous and highly competitive ASEAN economic region in which there is a free flow of goods, services and investments, a freer flow of capital, equitable economic development and reduced poverty and socio-economic disparities" (ASEAN Vision 2020).

The AEC is built on the current initiatives under the ASEAN Free Trade Agreement, the ASEAN Framework Agreement on Services, and the ASEAN Investment Area, with clear timelines specified for removal of nontariff barriers, harmonization of product standards and technical regulations, and conclusion of mutual recognition arrangements for priority sectors.

Prospects for ASEAN Economic Cooperation ASEAN faces a number of challenges in its journey toward greater economic integration. One major challenge is the widening economic gap within the grouping as ASEAN membership increased to 10 countries with significantly diverse economies and political regimes, and as the grouping is establishing more extra-ASEAN free trade agreements. The widening gap and the proliferation of extra-ASEAN free trade agreements, if not managed correctly, could lead to the weakening or marginalization of the ASEAN integration and, worse, to the irrelevance of ASEAN to the individual member countries' economic development.

See also Asia Pacific Economic Cooperation (APEC); free trade area; regionalism

FURTHER READING

ASEAN Secretariat. 1997. *ASEAN Economic Cooperation: Transition and Transformation*. Singapore: Institute of Southeast Asian Studies. A thorough review of the various schemes of ASEAN cooperation from the early stages of ASEAN to the 1990s.

ASEAN Vision 2020 (ASEAN Declaration at 13th ASEAN Summit). ASEAN Secretariat Website. Available at http://www.aseansec.org. The text of the 1997 summit statement.

Sjoholm, F., and J. L. Tongzon, eds. 2005. *Institutional Change in Southeast Asia*. New York: Routledge. A useful reference on the role of institutions in the ASEAN countries' economic development with a critical examination of key ASEAN institutions and their contributions to ASEAN economic performance, particularly during the 1997 98 Asian financial crisis.

Tan, G. 1996. *ASEAN Economic Development and Co operation*. Singapore: Times Academic Press. An overview of the economic development experiences of the ASEAN countries up to the 1990s and the challenges faced in their process of economic development.

Tay, S., et al. 2000. *A New ASEAN in a New Millennium*. Jakarta: Center for Strategic and International Studies and Singapore Institute of International Affairs. A valuable reference on ASEAN as an institution with a critical analysis of ASEAN's major political achievements, governing principles, and future challenges.

Tongzon, Jose L. 2002. *The Economies of Southeast Asia: Before and after the Crisis*. Cheltenham, UK: Edward Elgar. An in depth analysis of the economic development experiences of the ASEAN countries in the context of ASEAN economic cooperation before and after the Asian financial crisis of 1997 98.

. 2005. "The Role of AFTA in an ASEAN Economic Community." In *Road Map to an ASEAN Economic Community*, edited by D. Hew. Singapore: Institute of Southeast Asian Studies, 127 47. An in depth assessment of the role of AFTA in the building up of an ASEAN Economic Community, including the progress made in trade liberalization and the challenges faced in the next level of economic integration.

JOSE L. TONGZON

■ **asymmetric information**

An asymmetric information problem exists in a market if it is costly for some parties to observe the

characteristics or behavior of other parties, and an inefficient outcome results. One problem that arises due to asymmetric information is moral hazard, which exists when an arrangement that relieves a party of some risk causes the party to engage in riskier behavior. In credit markets, for example, use of a loan to finance a project means that the gains to the borrower are reduced if the project succeeds, because the principal must be repaid with interest. If the project fails, however, the creditor absorbs the loss of principal and interest, net of any collateral provided by the borrower. Risks are thus shifted from the borrower to the creditor, much as would occur with an insurance contract. If the creditor could costlessly observe how the borrower used the loan, then it might be possible to give the borrower contractual incentives to work hard and avoid risky projects, in order to hold down the risk of default on the loan. For example, future disbursements of the loan could be conditioned on the borrower's using the loan responsibly. However, if it is costly to observe the borrower's behavior, then moral hazard may arise, particularly if the loan is not fully collateralized or it is costly to take legal action against a borrower in default.

Adverse selection occurs in a market if it is difficult to distinguish parties that are good risks from those that are bad risks, and the parties that represent bad risks are particularly attracted to the market. For example, consider a set of borrowers that wish to finance projects that are identical in all respects, including offering identical expected returns, except that the projects vary in their riskiness. It can be shown that borrowers with riskier projects will have higher expected profits than other borrowers, and thus will have stronger incentives to borrow, while the expected rate of return to banks from these borrowers is lower than it is from other borrowers (Stiglitz and Weiss 1981). Thus the borrowers that tend to select themselves into the market cause problems for it.

If it were obvious which parties were the bad risks, those parties could be offered more stringent terms or even avoided altogether. Absent such information, in markets subject to adverse selection, the normal function that prices play in determining who participates in the market can result in a perverse outcome. In credit markets, higher interest rates limit who will want to borrow by making borrowing more burdensome, but those with the least risky projects will be the ones driven out of the market first, all else being equal. Those who are less likely to pay back the loan will be less concerned about paying higher interest. Because the average quality of the loans actually made can then deteriorate, the adverse selection problem can be worsened. A higher interest rate will also exacerbate the moral hazard problem for a given borrower, because the higher interest rate is burdensome only in the event that the borrower actually repays the loan.

Given these drawbacks of raising interest rates to limit access to credit, banks may choose to ration credit, so that borrowers are not allowed to borrow as much as they want at a given interest rate (Stiglitz and Weiss 1981). It is also possible that credit rationing may not occur, in which case a small increase in interest rates can lead to collapse of the loan market (Mankiw 1986).

Asymmetric Information and Financial Crises
Problems related to asymmetric information can contribute to, and be exacerbated by, financial crises, such as those experienced by South Korea, Thailand, and Indonesia in 1997 98. Financial crises that lead to bank failures can have real economic costs because of the loss of the customer relationships that these banks had cultivated. In the course of repeated interactions with their loan customers, banks gain a considerable amount of information on the customers' creditworthiness and other characteristics. This information is lost in a bank failure, which can worsen the moral hazard and adverse selection problems (Mishkin 1999). Thus financial intermediation becomes less effective, and real economic activity can be harmed.

Currency depreciation can exacerbate a banking crisis because it tends to worsen the balance sheets of individuals and corporations within the country that have borrowed in foreign currencies. This reduces the value of the collateral that can be obtained from these borrowers in the event of default and thus exacerbates

the moral hazard problem inherent in credit markets: borrowers with less to lose have weaker incentives to avoid default. Moreover, problems related to asymmetric information also limit how countries can respond to financial crises. For example, the higher interest rates required to combat inflation or stabilize the currency can aggravate adverse selection and moral hazard, as discussed earlier.

Information asymmetries can also exacerbate a tendency toward investor panic manifested by bank runs, capital flight, or asset price downturns, and can contribute to financial contagion between unrelated financial institutions or countries during crises.

Models of these phenomena typically posit that uninformed investors try to infer how asset prices will move based on the information revealed by these prices. Informed investors directly observe the fundamental determinants of asset prices, but the prices are also influenced by exogenous random factors. Moreover, informed investors are unable to drive asset prices directly to their equilibrium values dictated by the fundamentals because of various impediments to asset trading (such as trading costs, short-sale constraints, or borrowing constraints).

Now consider a negative shock to an asset price. Knowing that informed investors are limited in their ability to make trades, uninformed investors are uncertain whether the fundamental asset value is even lower, and therefore demand an additional risk premium to hold the risky asset, which forces its price down further. Thus asymmetric information exacerbates asset price movement.

For contagion between markets to occur, the simplest scenario is that the fundamental determinants of asset prices in two markets are correlated. Informed investors observe a shock to the fundamentals in one market, and in response reallocate their portfolios in the market that has experienced the shock and in other markets. The uninformed investor is uncertain whether asset price movements in either market are due to transactions by informed investors or due to random noise; given the greater uncertainties, uninformed investors demand a larger risk premium in both markets, which causes asset prices to move together.

Borrowing constraints can also trigger contagion, as shrinkage of investor assets in one country means that investors may be forced to sell their assets in other countries, as collateral requirements become binding (Yuan 2005). This contagion will also be exacerbated by the actions of uninformed investors. It will be stronger during asset price downturns than upturns and does not require that macroeconomic fundamentals in the countries be correlated. These findings are generally consistent with empirical evidence to date. Thus this model offers one explanation of why asset price shocks spread among East Asia, Latin America, and other parts of the world in 1998.

Lender of Last Resort Moral hazard becomes a policy issue in credit markets because of the lender-of-last-resort role played by various institutions. For example, central banks and other financial regulators are concerned about the possibility of a run on banks, which in turn could arise in part because it is difficult to distinguish solvent banks from insolvent ones. In order to lessen the risk of a run on banks, the central bank or other governmental institution usually provides implicit or explicit insurance to bank depositors. Bank liabilities may even be fully guaranteed on the grounds that the banks are too big or too politically important to be allowed to fail. The consequential moral hazard can give banks little reason to be cautious in their lending, or depositors little incentive to take the trouble to examine the financial soundness of one bank versus others. These problems can set the stage for a banking crisis.

Whether it is necessary or desirable for an international lender of last resort to assist countries in financial crises is a matter of some controversy. An argument in favor of such an institution is that the central bank of a country in crisis may have a limited capacity to restore the economy to health. For example, the provision of extra liquidity through monetary expansion could lead to currency depreciation and increases in inflationary expectations and interest rates (Mishkin 1999). Thus there may be an argument for an institution such as the International Monetary Fund (IMF) to lend foreign exchange to a country so that it can pay for imported inputs and

intervene to support its currency, say, to alleviate the real economic harms that a financial crisis can cause. As critics have observed, however, IMF bailouts of countries in financial distress can also cause moral hazard by relieving the governments of those countries of the painful effects of the inadequacies of their economic policies and by relieving international investors of the adverse consequences of investing in inherently risky environments such as emerging markets.

Following the IMF decision not to bail out Russia in 1998, the amount by which interest rates on bonds in developing countries exceeded those in industrial countries increased, particularly for countries with weaker economic fundamentals, indicating that investors were increasingly taking into account the risks of lending to developing countries (Dell'Ariccia et al. 2002). It cannot be concluded, however, that this change was for the better, even if it indicated that moral hazard was reduced: IMF lending that reduces risks could be a good thing on balance, and its positive effects would also be reflected in interest differentials across countries. For example, an IMF intervention in a developing country could alleviate market failures such as coordination problems among creditors to the country, by acting as a catalyst that restores confidence and induces creditors to resume lending rather than waiting for other creditors to take the first step.

Measures to Limit Moral Hazard At all the levels at which moral hazard exists in financial markets, measures can be taken to lessen its impact. Individual loan contracts can include provisions that make it harder for the borrower to take excessive risks or to shirk repayment of the loan. For example, loans can be disbursed in increments, conditional on performance to date. One reason for the existence of financial intermediaries like banks is that they can perform these disciplining functions at lower cost than can individuals, such as by threatening to withdraw future business from recalcitrant borrowers.

At the level of the banking system, the dilemma for policymakers is how to contain moral hazard and yet prevent or at least mitigate financial crises. Most

governments provide some form of implicit or explicit insurance for bank deposits. New Zealand does not, on the grounds that it eliminates depositors' incentives to monitor the riskiness of their banks, and because providing the deposit insurance at the same price to all banks implicitly subsidizes banks with riskier loan portfolios at the expense of those with more responsible ones. New Zealand instead tries to ensure that the public has complete and accurate information on banks' financial conditions. Some countries that provide deposit insurance, such as the United States, seek to lessen moral hazard by limiting deposit insurance to relatively small deposits, in order to provide larger depositors or other bank creditors with stronger incentives to monitor and price the risks embodied in banks' loan portfolios. In principle, the moral hazard from deposit guarantees can also be reduced through appropriate ongoing bank supervision and regulation. In practice, in many developing countries in particular, bank regulation remains problematic.

One approach to lessening the economic damage in the event of a crisis is for a lender of last resort to inject capital into financial markets. There may be merit in providing liquidity to the market in general rather than to particular distressed institutions, so that market forces can ultimately determine which institutions survive. If individual financial institutions are to be provided credit by the lender of last resort, Bagehot (1873) suggested that the terms should not be too attractive: the rate of interest should be higher than during normal times. Limiting moral hazard in this way is not without its own costs. Imposing a higher penalty rate can weaken the condition of the bank, signal to the market that the bank is in trouble, or induce bank managers to pursue a riskier strategy. However, Bagehot allowed that credit should be provided on collateral that would be marketable in normal times, so that a crisis-induced collapse of asset prices that caused collateral to shrink would not in itself limit the amount that could be lent.

If the crisis is due to a coordination failure, such as a panic-induced run on a bank, a lender of last resort may be able to suspend the obligations of the bank

temporarily, in a way that the bank could not do on its own due to credibility problems. This approach would not diminish the risk to which creditors of the bank are exposed, and thus could limit moral hazard. The lender of last resort could also coordinate a private bailout of the bank, in an effort to limit moral hazard. If the coordination of private institutions is at all coercive, however, it will in effect provide a subsidy to the distressed institution, in which case moral hazard reappears in an alternative form. Indeed, the bailout of the Long-Term Capital Management (LTCM) hedge fund in 1998 raised questions about whether private investors were coerced by the central bank. The LTCM bailout under the guidance of the Federal Reserve has been criticized for causing moral hazard.

Parallel issues have arisen for the IMF as an international lender of last resort that has bailed out countries in financial crisis. For example, IMF loans of foreign exchange to countries in crisis are disbursed over time, and the institution monitors whether the recipient country is complying with the performance and policy conditions negotiated as part of the loan agreement.

To avoid the moral hazard that IMF bailouts can cause, an alternative approach is to rely more on private-sector solutions. Along these lines, Mexico, Brazil, and other countries have introduced collective action clauses into their sovereign bond contracts since 2003. Such clauses are intended to facilitate debt restructuring in the event of a crisis, primarily by making it harder for a small minority of bondholders to block debt restructurings endorsed by a large majority. Inclusion of these clauses appears to have lowered borrowing costs for these countries, despite concerns that the clauses might make it easier for some countries, particularly those that are less creditworthy, to avoid repaying their debts.

A degree of deliberate ambiguity in the actions of the lender of last resort may also limit moral hazard. If it remains unclear whether the lender of last resort will provide a bailout in all situations, the parties who might or might not be bailed out will have incentives to act more responsibly. The decision

by the IMF not to bail out Russia in 1998, on the heels of its bailouts of South Korea, Thailand, and Indonesia in 1997 98, could be seen as creating such ambiguity.

A final restraint on moral hazard is for managers and owners of failed institutions to be punished: managers should lose their jobs and shareholders their capital. The argument may be applicable to national economies in crisis as well: a change of government tends to have a salutary effect in the recovery from a financial crisis, particular if mismanagement by the government was partly responsible for the crisis.

See also bail-ins; bailouts; balance sheet approach/effects; banking crisis; contagion; currency crisis; deposit insurance; financial crisis; International Monetary Fund (IMF); International Monetary Fund conditionality; International Monetary Fund surveillance; lender of last resort; spillovers

FURTHER READING

Bagehot, Walter. 1873. *Lombard Street: A Description of the Money Market*. Reprint, 1919. London: J. Murray. An early observer of financial markets, on how a lender of last resort should operate.

Dell'Ariccia, Giovanni, Isabel Schnabel, and Jeromin Zettelmeyer. 2002. "Moral Hazard and International Crisis Lending: A Test." IMF Working Paper WP/02/181. Washington, DC: International Monetary Fund. Examines whether IMF decisions to intervene or not to intervene in a crisis can affect borrowing costs for developing countries.

Eichengreen, Barry, and Ashoka Mody. 2001. "Bail ins, Bailouts, and Borrowing Costs." *International Monetary Fund Staff Papers* 47 (Special Issue): 155 87. Proposes measures to reduce moral hazard caused by IMF interventions through changes in international financial arrangements.

Mankiw, N. Gregory. 1986. "The Allocation of Credit and Financial Collapse." *Quarterly Journal of Economics* 101 (August): 455 70. Shows that adverse selection can cause a loan market to collapse.

Mishkin, Frederic S. 1999. "Global Financial Instability: Framework, Events, Issues." *Journal of Economic Perspectives* 13 (autumn): 3 20. Surveys international

financial crises of recent years; includes a useful discussion of how moral hazard and adverse selection may be exacerbated during a crisis.

Stiglitz, Joseph E., and Andrew Weiss. 1981. "Credit Rationing in Markets with Imperfect Information." *American Economic Review* 71 (June): 393 410. A seminal analysis of information problems in bank loan markets, which includes an argument that collateral requirements can cause adverse selection.

Yuan, Kathy. 2005. "Asymmetric Price Movements and Borrowing Constraints: A Rational Expectations Equilibrium Model of Crises, Contagion, and Confusion." *Journal of Finance* 60 (February): 379 411. Offers a theory of how asymmetric information can contribute to contagion across countries during financial crises and provides a useful survey of related literature.

STEPHEN MARKS

■ bail-ins

A bail-in is an agreement by creditors to roll over their short-term claims or to engage in a formal debt restructuring with a troubled country. Bail-ins are usually done in conjunction with a broader program of International Monetary Fund (IMF) lending and policy changes to help restore a country's economic and financial health. In a bail-in, creditors with claims coming due are asked to defer repayment deadlines and in some cases to agree to reduce their claims. In a bailout, by contrast, a country borrows hard currency reserves from the IMF, enabling it to pay off its maturing debt.

Academic models usually try to posit a clear choice between lender-of-last-resort financing and a standstill on all payments. In practice, though, policymakers rarely confront a binary choice between a complete standstill and lender-of-last-resort financing. Sources of existing or potential financial pressure on a crisis country are usually diverse: it is possible to bail in some creditors and bail out others. A strategy that combines a bailout (of some) and bail-in (of others) consequently can make sense. A successful bail-in of some creditors eliminates one potential source of financial strain. The benefits of securing additional financing from a set of private creditors, however, have to be balanced against the risk that attempting to bail-in some creditors will only prompt other creditors to run.

A bail-in is usually initiated when a country in financial trouble asks a set of its creditors to agree to roll over or reschedule their maturing claims. The debts coming due can be the obligations of the crisis country's government — for example, a maturing international sovereign bond — or they can be obligations of private borrowers (most often cross-border loans to the country's banks) in the crisis country. Convincing the country's creditors to defer payments, whether through a bond exchange or an agreement to roll over maturing bank loans, requires at least the implicit threat that the country will halt payments if the creditors do not agree on a restructuring. But the nature of the country's negotiations with its creditors can nonetheless vary. A country that tries to negotiate an agreement with its creditors to defer payments is acting quite differently from a country that just stops payments and demands that its creditors agree to reduce their claims.

Debt Restructuring and Burden Sharing Debt restructurings are a part of borrowing and lending; they would occur in the absence of any official-sector intervention. The trigger for a bail-in, however, is often a policy decision by the official sector not to provide a country with sufficient emergency financing to allow it to avoid a debt restructuring. The IMF, the Group of Seven (G7), and others can condition their lending on a requirement that a country keep its foreign exchange reserves above a designated level (effectively requiring the country to initiate a restructuring if it cannot find private sources of financing), refuse to lend in the absence of a restructuring, offer to provide financing to facilitate a consensual restructuring, or even help the country organize a rollover of its maturing loans. The Paris Club — a group of bilateral lenders — can also condition its willingness to restructure the debts a

country owes to bilateral creditors on a country's willingness to seek a comparable restructuring of the debt the country's government owes to private creditors. (The official sector generally has leverage over the debtor, not the debtor's creditors. The official sector sometimes can, however, exert leverage directly on certain types of creditors particularly banks. The official sector usually stops short of telling banks what to do. Nonetheless, major governments can make clear that it is in the banks' collective interest to cooperate to avoid default by agreeing to roll over their exposures.)

The line between a normal, voluntary market transaction issuing a new bond, a voluntary debt exchange and an involuntary concession to avert a crisis is not always clear. Some IMF programs rely on the expectation that the combination of financial support and policy adjustment will catalyze new private financial flows. Some debt exchanges done in the context of an IMF program occur at market rates and are altogether voluntary the creditors who do not participate can expect to be treated as well as creditors who do participate. Because these exchanges must be done at market rates, they are often expensive. The most famous example is Argentina's megaswap in the summer of 2001: the swap extended the maturity of Argentina's bonds, but at an implied annual interest rate of close to 15 percent. This transaction allowed Argentina to avoid default only for six months, however. These voluntary exchanges are not a true bail-in: private creditors are extending credit at a market rate in the expectation of earning a commercial profit, not making a concession to help the country through a crisis.

Other debt exchanges are done at below-market rates to avoid an imminent default. These transactions are also voluntary in some loose sense. Creditors often are willing to defer payments at an interest rate lower than the prevailing market interest rate at the time of the exchange as in Uruguay's 2003 exchange or even accept a deep "haircut" as in Argentina's 2005 exchange rather than hold debt that the sovereign debtor is not willing (or able) to pay. A haircut typically involves a reduction in the

face value, a reduction in the coupon (the amount to be paid at fixed intervals), or a reduction in both the face value and the coupon of the bond. The actual losses experienced by creditors who mark to market (that is, value the bond as an asset at its open-market price) depend on the price at which they bought the bond and the discount rate the market assigns to the payment stream on the restructured bonds. The decision to voluntarily agree to these kind of terms reflects the fact that the alternative to a restructuring is often default and creditors lack the legal ability to force a sovereign debtor that is in default to resume payments.

Throughout the 1990s, the G7 and the IMF had trouble reaching agreement on the right term to use to describe efforts to secure the coordinated provision of emergency financing from a country's existing private creditors. Calls for more "burden sharing" were considered too heavy handed: no private creditor happily takes on a burden. Talk of "constructive engagement" with private creditors was considered a bit too diplomatic: private creditors generally preferred other forms of engagement. The relatively informal term *bail-in* drew attention to the bailouts that sometimes were provided to avoid bail-ins. The most widely used term *private-sector involvement in crisis resolution* suited the international bureaucracy well: it was easy to reduce to an acronym (PSI).

Acronyms can still generate impassioned debate. "Market fundamentalists" opposed all forms of official intervention. They wanted to scale back IMF lending. But they also opposed the official sector's efforts to facilitate the coordinated provision of emergency financing by private creditors. This group wanted both fewer bailouts and fewer officially organized bail-ins. A more pragmatic group including many in the G7 and the IMF hoped to combine official financing with attempts to involve private creditors. This group argued that commitments by prominent creditors not to take their money out would help to limit the distortions introduced by official crisis support. Many Europeans viewed efforts to involve private creditors as a direct substitute for large-scale official financing they

wanted more bail-ins to reduce the need for financial bailouts. They emphasized the need to change the institutions for debt restructuring in order to make restructurings less disruptive whether through the introduction of collective action clauses in international sovereign bonds or the development of an international bankruptcy regime. Many emerging economies sometimes with support from those in the U.S. Treasury who believed recent emerging market crises were overwhelmingly the product of an international analogue to bank runs wanted to banish all talk of combining bailouts with bail-ins, arguing that any effort to bail in some groups of creditors would scare market participants and keep IMF financial support from generating the desired improvement in creditor confidence.

Bail-ins were rarely part of the official sector's initial response to market turmoil. Most countries approached their private creditors to seek emergency financing only after an initial (and sometimes limited) round of official financing failed to end their financial trouble. There have, nonetheless, been important successes. Korea convinced the international banks that had lent to Korean banks first not to demand payment on their maturing loans and then to reschedule these loans. The rollover agreement eliminated the immediate threat of default and along with continued financing from the IMF helped to pave the way for Korea's financial recovery. Uruguay successfully combined a very large credit line from the IMF with a bond restructuring. The credit line stopped Uruguay's bank run and provided Uruguay the time needed to execute the bond exchange; the exchange assured that the IMF's funds were not used to finance a reduction in the country's bond exposure. Pakistan and the Dominican Republic also restructured their international bonds without stopping payments.

Other bond restructurings came only after the country had fallen into general default. Such restructurings sought to clean up an existing financial mess rather than avert a deeper financial crisis. Russia's 2000 restructuring of its "London Club" debt was technically a restructuring of syndicated bank

loans, but since most of these loans had been securitized and sold into the market, it resembled a bond exchange. Both the Ukraine and Ecuador restructured their international sovereign bonds in 2000. Argentina's restructuring, though, dwarfs the others in size and complexity: in 2005 Argentina sought to restructure 152 separate bond issues with a face value of more than $80 billion.

The legal challenges associated with a bond restructuring, though significant, have to date proved to be smaller than many initially feared. The difficulty in keeping a bond restructuring from leading to broader financial collapse and specifically a domestic bank run proved larger than expected, however. In practice, many "international" sovereign bonds (bonds governed by a foreign law) are held not by international investors, but by the domestic banking system.

Crisis Management: Finding the Right Balance
No single measure can gauge the success of efforts to obtain crisis financing from the country's private creditors. Success requires convincing private creditors to contribute, whether by deferring payments or by agreeing to reduce their claims on the crisis country. But success also requires that the private creditors' contribution not come at the expense of other goals including preventing a sharp fall in output or triggering a broader run that leaves the country in a deep financial hole. Finding strategies that strike the right balance between these sometimes conflicting goals has been a constant challenge.

See also bailouts; banking crisis; currency crisis; financial crisis; international financial architecture; International Monetary Fund (IMF); International Monetary Fund conditionality; International Monetary Fund surveillance; Latin American debt crisis; lender of last resort

FURTHER READING
Blustein, Paul. 2005. *And the Money Kept Rolling In (and Out): Wall Street, the IMF, and the Bankrupting of Argentina*. New York: Public Affairs. The best account of Argentina's crisis.
Eichengreen, Barry. 2001. "Managing Financial Crises." Felix Neubergh Prize Lecture. Goteborg University, November 13. Downloadable from http://www.

econ.berkeley.edu/~eichengr/policy/crises nov01.pdf. A concise summary of the policy debate from an influential economist.

———. 2002. *Financial Crises and What to Do about Them.* New York: Oxford University Press. A book length examination of financial crises.

Haldane, Andrew, and Mark Kruger. 2001. "The Resolution of International Financial Crises: Private Finance and Public Funds." Bank of Canada Working Paper 2001 20 (November). Ottawa: Bank of Canada. An influential paper that made the case for using general payment standstills in a broader range of crises.

International Monetary Fund. 2000. Communiqué of the International Monetary and Financial Committee of the IMF Board of Governors (September 24). Downloadable from http://www.imf.org/external/np/cm/2000/092400.htm%20. The best example of the official sector's effort to reach consensus on how crises should be managed and to spell out how private creditors might be "involved" in a range of different circumstances.

Krueger, Anne. 2001. "A New Approach to Sovereign Debt Restructuring." Speech given at the National Economists' Club Annual Members' Dinner, American Enterprise Institute, Washington (November 26). Downloadable from http://www.imf.org/external/np/speeches/2001/112601.htm. Proposed the creation of a new "statutory" international bankruptcy regime.

Rogoff, Kenneth, and Jeromin Zettelmeyer. 2002. "Bankruptcy Procedures for Sovereigns: A History of Ideas, 1976 2001." *IMF Staff Papers* 49 (3): 471 507. Washington, DC: International Monetary Fund. Downloadable from http://www.imf.org/external/pubs/ft/staffp/2002/03/pdf/rogoff.pdf.

Roubini, Nouriel, and Brad Setser. 2004. *Bailouts or Bail ins.* Washington, DC: Institute for International Economics. Comprehensive survey of experience with debt restructuring; argues that bail ins can be successfully combined with bailouts in some circumstances.

Sturzenegger, Federico, and Jeromin Zettelmeyer. 2006. *Debt Defaults and Lessons from a Decade of Crises.* Cambridge, MA: MIT Press. An up to date analysis of experience with debt restructuring, including calculations of investor losses in recent restructurings.

BRAD SETSER

■ bailouts

Bailouts broadly refer to large loans from the official sector usually the International Monetary Fund (IMF) to a country facing difficulties repaying its maturing debts or needing to intervene in the foreign exchange market to defend its exchange rate in the face of the withdrawal of foreign investment or domestic capital flight. These loans are provided directly to the crisis country to augment its reserves in the face of acute balance of payments difficulties. Augmenting a crisis country's hard currency (international) reserves often allows the country to avoid what otherwise would have been an almost certain default on its private debts and, in some cases, to intervene in the currency market to maintain an exchange rate peg at a level that would otherwise be unsustainable. Consequently, large loans that help a country try to avoid economic and financial collapse also help external and domestic creditors with foreign-currency debts coming due, as well as those looking to exchange local currency for dollars, euros, or another currency.

The term *bailout* could be applied to all loans provided for balance of payments support by the official sector. In practice, the policy debate over these loans focused primarily on the largest of these loans, and specifically those loans whose size exceeded the IMF's traditional lending limits of 100 percent of quota in a year, or 300 percent of quota over three years. (A country's quota determines the size of a country's contribution to the IMF as well as its voting and borrowing rights.) The first prominent loan to breach these limits was provided in 1995 to Mexico. By the end of 2007, the IMF had provided "exceptional" levels of IMF financing to an additional eight emerging market economies facing acute balance of payments difficulties. Indeed, "exceptional" financing that is, financing in excess of the IMF's normal lending norms effectively became the new norm for large emerging market economies that encountered balance of payments difficulties.

Using the size of an IMF loan relative to quota to distinguish between small and often uncontroversial IMF loans and those exceptionally large IMF loans creates some difficulties. First, the size of a country's

IMF quota maps only imperfectly to other relevant economic criteria Korea in 1997 and Turkey in 2001 and 2002 got exceptionally large IMF loans relative to their quotas, in part because their quotas were quite small relative to their current gross domestic product (GDP). Second, in several prominent cases, the multilateral development banks and the Group of 10 countries (a group of economically advanced countries) provided large amounts of financing alongside the IMF, so looking only at IMF financing may understate the scale of official support. For example, the United States provided Mexico almost as much financing from its Exchange Stabilization Fund as the IMF and a group of bilateral creditors coordinated by the Bank for International Settlements (BIS) disbursed alongside the IMF in 1999. The World Bank and the Asian Development Bank pushed by the United States provided a large amount of emergency financing to Korea, and Japan provided additional bilateral funds to Thailand (the bilateral "second-line" commitments included in the initial packages for Korea and Indonesia were never disbursed). Loans that breached the IMF's own access limits also tended to be the largest loans when measured by other criteria. Those countries that got large amounts of financing from other official sources also tended to get large amounts of financing from the IMF.

How Does the IMF Differ from Domestic Lender of Last Resort? An IMF loan to a country is in some ways similar to emergency financing from a domestic lender of last resort (typically, the country's central bank). Both provide liquidity cash to those in need of it. Just as a loan from a domestic lender of last resort lets a troubled bank pay off its depositors, an IMF loan also lets some of the crisis country's creditors "off the hook" by financing the repayment of the debt. Just as support from a domestic lender of last resort can convince a troubled bank's creditors its depositors to give the bank time to try to work through its difficulties, international crisis financing can convince the country's creditors to give it time to work through its problems.

But there are also important differences between the IMF and a domestic lender of last resort. First, the IMF's lending capacity is constrained by the amounts members have placed on deposit with the fund. A domestic lender of last resort has no similar constraints, at least as long as it lends in the country's own currency. Second, IMF loans are usually tied to policy changes, and in order to encourage the country to implement those changes, IMF lending is rarely disbursed all at once; instead, IMF loans are usually disbursed in a series of tranches. A domestic lender of last resort can provide all necessary funds up front. Third, the IMF does not lend against collateral; rather, crisis countries, by long-standing convention, pay the IMF even if they are not paying their other creditors. This allows the IMF to lend to crisis countries at modest rates without taking losses. Finally, many domestic bank "bailouts" do more than just provide a cash-strapped domestic bank with emergency liquidity. They also typically require giving a troubled bank a government bond a new financial asset to prevent depositors (and sometimes even the shareholders) in the bad bank from taking financial losses. International bailouts, by contrast, do not increase the net assets of a troubled emerging economy. The country's external reserves rise, but so do its external debts. The international taxpayers who put up the money needed to make an IMF loan expect to get repaid in full.

Consequently the term *bailout* is perhaps too negative a term for emergency crisis financing, as it suggests that the crisis lender is picking up losses that otherwise would have been borne by the country and its creditors. The term *rescue loan* is probably too positive, however, as not all "rescues" have succeeded. Most neutral terms such as *large-scale official crisis lending* sound bureaucratic. The lack of an agreed term itself may be indicative of the ongoing debate about the wisdom of large-scale lending to crisis countries.

Debate Surrounding IMF Crisis Lending Proponents of large-scale IMF lending argue that financial integration is generally beneficial. But they also recognize that it can increase the risk that a country with correctable policy problems can be forced into a disruptive default by a self-fulfilling crisis of confidence. Concerns that other external creditors will not

roll over their short-term debts as they come due can lead all external creditors to demand payment as soon as possible. A country's own citizens can also decide that they want to shift their savings abroad before the country runs out of reserves. Such a shift from domestic to foreign assets puts enormous pressure on the country's reserves if it has a fixed exchange rate regime, on its exchange rate if it has a floating exchange rate regime, or on both reserves and the exchange rate if it has a managed float. Sergei Dubinin, the chairman of Russia's central bank during the 1998 crisis, observed, "We can play games against the market, against the banks even, but we can't do anything if the entire population wants to change rubles into dollars" (Blustein 2001, 266–67).

Such crises of confidence have usually stemmed at least in part from doubts about a country's ability to put in place needed corrective policies, not just doubts about the willingness of other creditors to maintain their exposure. But as the run intensifies, a country can be pushed toward default well before it has time to show whether it can make the policy changes to ensure its long-run solvency.

Critics of large-scale lending have raised a host of objections. Some object to all official intervention in private markets. Official action that insulates a country and its creditors from paying the full price for their mistakes only encourages more bad policies and additional risk-taking—so called moral hazard. Others object to the conditions attached to IMF loans,

Table 1
How quickly were IMF (and bilateral first line) loans disbursed, and how fast were they repaid?

	Peak disbursement, $ billion (% of GDP)	Quarters to reach peak	Quarters to repay ½ peak disbursement	External debt precrisis (% of GDP)	Fiscal debt precrisis (% of GDP)
Mexico	27.6 (6.8)	4	9	34%	31%
Thailand	11.2 (6.2)	12	18[a]	60%	5%
Indonesia	10.8 (4.7)	13	35	43%	24%
Korea	19.4 (3.7)	4	8	32%	12%
Russia	5.1 (1.2)	2	4	35%	52%
Brazil (98–99)	17.5 (2.2)	3	7	25%	40%
Turkey (00–02)	23.2 (12.6)	13	24	57%	56%
Argentina	14.6 (5.1)	4	22	51%	45%
Uruguay	2.7 (14.6)	11	19	81%	38%
Brazil (01–02)	33.3 (5.7)	10	18	44%	65%

Sources: IMF and bilateral first line lending data from IMF, U.S. Treasury; debt data from Moody's (apart from Mexico's precrisis debt data, which is from the IMF). Moody's debt numbers for Brazil are higher than other sources. The IMF has Brazil's 1997 (precrisis) debt to GDP at 35% rather than 40%; and Brazil's 2000 (precrisis) debt to GDP at 49% rather 65%.

[a] Thailand's IMF exposure peaked after 9 quarters, and it repaid half of that exposure after 17 quarters. At that point in time, it had not repaid ½ its bilateral lending. However, we do not have data indicating Thailand's bilateral repayments after the end of 2001.

whether fiscal austerity, monetary tightening, or domestic financial and corporate reform. This criticism was particularly vociferous after the Asian crisis; many argued that the IMF's conditions themselves contributed to the cascading loss of confidence. Many argued that the large bailouts of the 1990s and the first few years of the new millennium were a marked change from the 1980s, when the typical IMF loan was smaller (in relation to quota, though quotas were larger in relation to GDP) and was combined with commitments by the country's largest creditors   the commercial banks   to roll over their claims.

Finally, some argued that the real problem was not that the IMF was lending too much, but rather that the IMF was not able to lend enough. The IMF risked providing enough money to allow a lucky few to exit, but not enough to assure that the run would stop and the country would have time to put in place corrective policies.

This highlights a key point: in practice, even the largest IMF loans typically fell short of providing sufficient funds to cover all potential drains on the country's hard currency liquidity. For example, Mexico received enough money to cover payments on the government's maturing short-term dollar-linked debt, the famous *tesobonos*, but not enough to cover all potential sources of capital flight. Even a loan large enough to cover the most obvious sources of payment difficulties works only if additional sources of financial pressure do not also materialize.

Has IMF Crisis Lending Been Successful? Several large IMF bailouts achieved most of their intended goals. Mexico, Korea, and Brazil (in 1999) all avoided default, recovered market access relatively quickly, and paid back their official creditors quite rapidly. Eight quarters after the onset of their crises, all had paid back more than one-half of their initial loans (see table 1). Brazil, however, is less obviously a success than Mexico and Korea, as the rise in its debt levels during the 1998 99 crisis laid the foundation for its 2002 crisis. Brazil did not pay its 2002 bailout loan back in eight quarters   in part because the large loan in 2002 came on top of an earlier loan in 2001. But it otherwise resembles the "success" stories: it fully repaid the IMF at the end of 2005, and its debt

levels were coming down through the end of 2007. Other rescues achieved their goals, but not as quickly. Thailand (1997), Turkey, and Uruguay all avoided default, regained market access, resumed growth, and eventually made significant payments back to the IMF. All these countries entered into their crises with higher debt levels and generally have taken longer to repay the IMF than Mexico, Korea, and Brazil (particularly in 1999).

Three bailouts clearly failed to achieve their initial goals, however: Russia, Argentina, and Indonesia   whether because the country failed to carry out its commitments to the IMF or because the IMF backed a flawed strategy. In 1998, Russia was cut off quite quickly, after one $5 billion disbursement from the IMF's new program   a decision prompted in part by the IMF's substantial exposure from its previous lending programs. Indonesia received only $4 billion in the early stages of its crisis. Most of the funds it received from the IMF came later on, after its corporate sector had fallen into general default. Argentina, in contrast, received almost $15 billion in an unsuccessful attempt to ward off default.

By the end of 2007, the IMF had not approved a new large loan   setting aside those loans extended to refinance existing IMF loans   since 2002. In retrospect, the period of large-scale IMF lending that followed Mexico's crisis may be viewed as facilitating the transition of many emerging economies from a point where they held too few reserves to navigate periods of financial volatility to a point where most emerging economies held more reserves than they needed. During the turbulent period between 1995 and 2003, the size of IMF lending was striking. By 2005, though, the size of even the largest IMF loans seemed fairly small relative to the reserves of the typical emerging economies. For all the criticisms leveled at the IMF, most emerging markets concluded that they needed to hold far more reserves than the IMF was ever willing to make available.

See also asymmetric information; bail-ins; banking crisis; capital flight; contagion; currency crisis; financial crisis; international financial architecture; international liquidity; International Monetary Fund (IMF); International Monetary

Fund conditionality; International Monetary Fund surveillance; international reserves; lender of last resort

FURTHER READING

Blustein, Paul. 2001. *The Chastening: Inside the Crisis that Rocked the Global Financial System and Humbled the IMF*. New York: Public Affairs. A readable and critical account of the IMF's response to the Asian financial crisis.

Fischer, Stanley. 1999. "On the Need for an International Lender of Last Resort." *Journal of Economic Perspectives* 13 (4): 85 104. The case for an international lender of last resort.

Jeanne, Olivier, and Jeromin Zettelmeyer. 2001. "International Bailouts, Moral Hazard, and Conditionality." *Economic Policy* 33: 409 32. Argues that IMF lending may sustain bad domestic policies since IMF lending unlike domestic bailouts is almost always repaid.

Ortiz, Guillermo. 2002. "Recent Emerging Market Crises What Have We Learned?" Per Jacobsson Lecture. Basel: Bank for International Settlements. Downloadable from http://www.perjacobsson.org/lectures/2002 ortiz.pdf. A sympathetic treatment of the case for large scale IMF lending.

Roubini, Nouriel, and Brad Setser. 2004. *Bailouts or Bail ins*. Washington, DC: Institute for International Economics. A comprehensive survey of recent experience with crisis resolution along with the academic debate.

Rubin, Robert E., and Jacob Weisberg. 2003. *In an Uncertain World: Tough Choices from Wall Street to Washington*. New York: Random House. There is no better account of the U.S. decision to bail out Mexico and the resulting policy debate.

BRAD SETSER

balance of payments

A balance of payments is an international accounting record. It records the economic transactions of the residents of one country with the residents of other countries during a fixed period (month, quarter, year). The payments are usually measured in a country's home currency, although the accounts are sometimes expressed in U.S. dollars to facilitate cross-country comparisons. The payments are broken down into categories goods, services, transfers, investments, and official reserves. Inflows and outflows in each category are measured and a net figure for each is calculated. Strictly speaking, each category has a balance and, as a matter of practice, several balances are regularly reported by government agencies and the financial press. *The balance of payments* is a generic term that refers to payment flows in some or all of the different categories.

Although balance of payments accounts are constructed using sophisticated accounting principles, including double-entry bookkeeping, a country's cross-border transactions are not measured with the accuracy of a corporation's accounts. Balance of payments figures are statistically estimated based on sampling data gathered by government agencies. In the United States, business firms are required by law to provide information on international transactions. The Office of Management and Budget, the Customs Service, the Department of Commerce, and the Treasury Department gather or compile such information. Although these agencies do their utmost to provide reliable estimates, of necessity the accounts contain errors; the statistical estimates are based on data that are incomplete and otherwise imperfect.

Uses Business firms, labor unions, academics, lawyers, investors, and government officials take an interest in the balance of payments. Firms base financial and competitive decisions on the balance of payments. Firms and labor unions use balance of payments figures to argue for favorable tariffs or subsidies. Lawyers use information from the balance of payments to initiate cases in national courts or before the dispute resolution bodies of the World Trade Organization (WTO). Economists develop principles to explain the movement of goods and investments across borders, and they recommend policies based on those principles. Government officials are responsible for the policies that govern cross-border economic transactions. Some government policies, such as the trade restrictions imposed during the Cold War, discourage international transactions, while other policies, such as free trade agreements, encourage them. The balance of pay-

ments also helps us to understand how the world trading system is evolving. The flow of foreign aid from the United States to Europe in the 1950s, the increase in the exchange of services in the 1980s, and the rise of foreign direct investment in the 1990s all show up in the balance of payments accounts.

History The interest in international transactions goes back to at least the 14th century. The mercantilists, a group of British writers, urged government policies that encouraged exports and discouraged imports. They argued that a surplus of exports over imports would preserve the country's specie (metallic money supply) and enhance the wealth of the country. Borrowing the idea of an accounting balance from the double-entry bookkeeping of Italian merchants, the mercantilists introduced the term *balance of trade* into economic discussion in the 1600s. In the 18th century, David Hume pointed out that a country cannot run a permanent trade surplus (his famous price-specie flow mechanism), and Adam Smith roundly criticized the mercantilists for conflating money and wealth. The balance of payments received a great deal of theoretical attention in Britain's 19th-century monetary debates; the effects of fiduciary media (paper money) on the trade balance were especially controversial. European and U.S. governments collected international trade statistics in the 19th and early 20th centuries, but the modern system of payments accounting was not established until the end of World War II, when agreements at Bretton Woods empowered the International Monetary Fund (IMF) to gather and organize international trade data. World leaders recognized that the breakdown of the international trading system during the 1930s had exacerbated the Great Depression. They hoped that systematic gathering and presentation of trade data would enhance international cooperation.

Accounts All international economic transactions are divided into two main accounts the current account and the awkwardly named capital and financial account.

Current Account The current account is broken down into four categories goods, services, income, and transfers. Goods are the tangible items agricultural products, manufactured goods, and commodities such as copper and crude oil that are traded across borders. For most countries, goods are the main component of the current account. In 2004, the United States exported slightly more than $800 billion in goods, about 50 percent of current account credits for the year. Japan exported $539 billion in goods, about 70 percent of its current account credits (see table 1). Services are the exchange of intangibles across borders. Travel, transportation, royalties and license fees, and financial (banking, investment, insurance) and military services are the main ones traded across borders. In 2004, the United States exported $340 billion in services, Japan $98 billion. Business and personal travel is an important component of services in many countries; $74 billion, or 21 percent, of U.S. service export revenue came from travel in 2004. Income in the current account is primarily the dividends and interest paid to residents of a country from their investments abroad. The income can result from a resident in one country holding securities in foreign companies or governments, or from interest earned on loans to residents of other countries. For a country that invests large sums abroad but discourages foreign investment into the country, inflows of income outweigh the outflows by a large margin in Japan's case, $113 billion to $27 billion in 2004. Besides income from investments abroad, the income account also includes compensation to temporary workers, such as seasonal farmworkers and short-term business consultants. Compared to investment income, employee compensation is small for most countries; in 2004, the United States, for example, paid $9 billion to temporary workers from other countries, but $340 billion in interest and dividends to foreign investors.

The final item in the current account is unilateral transfers. The main items in this account are foreign aid (government grants of cash, food, clothing, etc. to the residents of other countries), charity (private donations to residents of other countries), and worker remittances (money sent home by workers employed abroad). Residents of wealthy economies usually spend considerable sums on government aid and charitable activities aimed at helping residents of

Table 1
Analytic balance of payments, United States and Japan (billions of U.S. dollars)

Line		U.S. 2000	U.S. 2004	Japan 2000	Japan 2004
1	**A. Current Account**	**416.00**	**668.07**	**119.66**	**172.06**
2	Goods: exports, f.o.b.	774.63	811.03	459.51	539.00
3	Goods: imports, f.o.b.	1224.32	1472.96	342.80	406.87
4	*Trade Balance*	*449.78*	*661.93*	*116.72*	*132.13*
5	Services: credit	296.85	340.42	69.24	97.61
6	Services: debit	225.34	296.07	116.86	135.51
7	*Balance on goods and services*	*378.27*	*617.58*	*69.09*	*94.23*
8	Income: credit	350.92	379.53	97.20	113.33
9	Income: debit	329.86	349.09	36.80	27.63
10	*Balance on goods, services, & income*	*357.22*	*587.14*	*129.49*	*179.93*
11	Current transfers: credit	10.83	17.92	7.38	6.91
12	Current transfers: debit	69.61	98.85	17.21	14.78
13	*Balance on current account*	*416.00*	*668.07*	*119.66*	*172.06*
14	**B. Capital Account**	**0.93**	**1.65**	**9.26**	**4.79**
15	Capital account: credit	1.08	1.13	0.78	0.44
16	Capital account: debit	2.01	2.78	10.04	5.23
17	*Total: Groups A plus B*	*416.93*	*812.16*	*110.40*	*167.26*
18	**C. Financial Account**	**486.66**	**581.79**	**78.31**	**22.49**
19	Direct investment abroad	159.21	252.01	31.51	30.96
20	Direct investment in U.S.	321.27	106.83	8.23	7.80
21	Portfolio investment assets	127.91	102.38	83.36	173.77
22	Equity securities	106.71	83.20	19.72	31.47
23	Debt securities	21.19	19.18	63.64	142.30
24	Portfolio investment liabilities	436.57	762.70	47.39	196.72
25	Equity securities	193.60	61.91	1.29	98.22
26	Debt securities	242.97	700.79	48.67	98.44
27	Financial derivatives	. . .	. . .	4.67	2.41
28	Financial derivatives assets	. . .	. . .	106.74	56.44
29	Financial derivatives liabilities	. . .	. . .	111.41	54.06
30	Other investment assets	273.11	503.92	4.15	48.01
31	Monetary authorities	. . .	. . .	. . .	. . .
32	General government	0.94	1.22	1.89	3.87
33	Banks	133.38	356.13	36.51	3.24
34	Other sectors	138.78	149.00	38.77	55.12
35	Other investment liabilities	289.05	570.58	10.21	68.30
36	Monetary authorities	2.52	52.77	. . .	. . .
37	General government	0.39	1.39	0.93	0.98
38	Banks	122.72	392.96	28.22	42.73
39	Other sectors	169.24	123.46	37.49	24.59
40	**D. Net Errors and Omissions**	**69.44**	**85.13**	**16.87**	**28.90**
41	*Overall balance*	*0.29*	*2.80*	*48.95*	*160.85*
42	**E. Reserves and Related Items**	**0.29**	**2.80**	**48.95**	**160.85**
43	Reserve Assets	0.29	2.80	48.95	160.85
	Conversion rates: yen per U.S. dollar			*107.77*	*108.19*

Source: Adapted from *Balance of Payments Statistics Yearbook*, International Monetary Fund, 2005.

other countries. In 2004, the residents of the United States disbursed foreign aid of $30 billion, private charitable contributions of $39 billion, and worker remittances of $30 billion to other countries. Residents of developing countries are usually net recipients of aid, charity, and remittances.

Financial Account Like the current account, the financial account is divided into four main kinds of transactions. The first of these is direct investment. Whenever residents in one country make an investment abroad that gives them a "substantial voice" in management, the investment is considered direct. A firm may do this by investing in plant and equipment or by purchasing enough stock (10 percent or more of outstanding shares) to influence the managerial decisions of an existing firm. In 2004, U.S. investors spent $252 billion on direct investments abroad, $176 billion of which was reinvested earnings. Foreign investors spent $106 billion on direct investment in the United States, $56 billion of which was reinvested earnings.

The second kind of investment transaction is portfolio investment, which is the purchase of bonds and equities in one country by residents of another. Although equities are issued mainly by private firms, bonds are issued by both firms and governments, so portfolio investment includes the purchase of government bonds by residents of other countries. In 2004, foreign residents invested $62 billion in U.S. equities and $701 billion in U.S. bonds, $458 billion of which was government bonds. By way of contrast, foreign residents invested $197 billion in Japan. The investment was split evenly, with $98 billion going into equities and $98 billion into bonds, of which $53 billion was government bonds. The government absorbed 60 percent of portfolio investment in the United States, compared to 27 percent absorbed by the Japanese government.

A third category of investment is a residual, named "other" in the accounts, which includes trade credit, bank loans and deposits, and currency exchanges. In 2004, foreign residents increased their residual investments in the United States by $570 billion, $407 billion of which represented an increase in the holdings of U.S. currency and bank deposits.

In Japan, the pattern was rather different. Foreign residents increased borrowing from Japanese firms and government agencies by $81 billion but reduced their holdings of the yen by $13 billion, so residual investment increased by $68 billion.

The fourth category of investment in the financial account is reserve assets, which are funds available to monetary authorities to finance payment deficits. Central banks and treasuries use gold, holdings with the IMF, and foreign exchange (currencies and deposits) for such financing. In 2004, monetary authorities in the United States financed an overall deficit of $2.8 billion, mainly by altering its reserve position within the IMF. Japan had an overall surplus of $161 billion, with which it acquired foreign exchange (i.e., international reserves).

Double-Entry Bookkeeping Every economic transaction entered into a country's balance of payments accounts contains both a debit and a credit. On the current account, debits and credits are apparent. When a country exports, the account shows a credit; when a country imports, the account shows a debit. On the capital account, the debits and credits can be somewhat confusing. When a company receives payment for its exports, it is recorded as a debit in the capital account; when a company makes a payment for an import, it shows up as a credit in the capital account. One way to make sense of credits and debits is to view them as sources and uses of funds, where *funds* means financial assets used in international transactions. Currencies, deposits in banks, trade credits and other forms of loans are examples of funds. In the balance of payments accounts, a source of funds is a credit; a use of funds is a debit. When a firm sells a bond to a foreign resident, the sale will show up as a credit (a source of funds) in the capital account of the home country. By selling a bond to a foreign resident, a firm has also increased its liabilities. An increase in liabilities is a source of funds and therefore a credit in the capital accounts. Conversely, when a firm redeems a bond, it reduces its liabilities by using funds. A decrease in liabilities is therefore a debit in the capital account. The opposite is true of assets. When a resident of one country purchases equity shares in a foreign firm, she uses funds; an

increase in assets appears as a debit in the financial account. When this investor sells her shares, she decreases her assets and increases her funds. A decrease in assets appears as a credit in the capital accounts.

Balances If we accurately record all of the flows of payments between countries, they will always balance. Yet policymakers and others are often concerned with payment imbalances. Imbalances occur when we consider only particular types of transactions within the accounts, rather than all transactions. For example, if we consider only merchandise transactions, the resulting balance is officially called the balance on goods and is reported in the press as the trade balance. The trade balance for the United States, shown in line 4, was a negative $662 billion in 2004, while Japan's was a positive $132 billion. Although goods are usually the biggest item in a country's current account, services have become increasingly important. In 2004, the United States exported $340 billion and imported $296 billion in services. The United States' balance in services was a positive $44 billion, while Japan's was a negative $38 billion. The press does not usually report the services balance by itself, however. Instead, it reports the balance on goods and services, which for 2004 was negative $618 billion for the United States and positive $94 billion for Japan. This balance is particularly important, because it enters the national income accounts of each nation.

The gross domestic product of a country, measured in expenditures, equals consumption plus investment plus government expenditures plus exports minus imports, or $Y = C + I + G + (X - M)$. In this equation, $X - M$ is the balance on goods and services. Since U.S. residents spent more on imports than foreign residents spent on U.S. exports in 2004, the balance on goods and services subtracted from the gross domestic output of the United States. If we add to the balance on goods and services the balance on income and the balance on unilateral transfers, neither of which is widely reported, we get the current account balance, which is closely watched by policymakers. If this balance is negative, it means that a country's current transactions are being financed by investment from abroad (or from official reserves).

There is nothing wrong or even worrisome in this, as long as the investment is being used for productive purposes. If the investment does not produce a higher future income, however, then its repayment will lower a country's future standard of living.

Another balance closely watched is the overall balance. This is the sum of all current and financial account transactions, adjusted for errors and omissions, except for official reserve transactions, and is shown on line 41 of table 1. A negative overall balance means that a country is balancing its payments by draining its official reserves. If a country does not balance its current account deficit with a financial account surplus, or if it does not balance a financial account deficit with a current account surplus, then it must make up the difference by using its accumulated reserves. Governments will often use their reserves to balance the overall account over short periods, but they can do so only until their reserves are depleted. A large and persistent deficit in the overall balance indicates that a country's trade and financial policies are not sustainable.

The Effects of Monetary and Fiscal Policies on the Balance of Payments A country's balance of payments results from a large number of diverse transactions, conditions, events, and policies. Natural resources (copper in Zambia, oil in Saudi Arabia), comparative advantage (automobiles in Japan), entrepreneurial discovery (the computer industry in the United States), as well as exchange rates, interest rates, income levels and growth, savings rates, political stability, wars, tariffs, and subsidies all affect a country's balance of payments. Moreover, these influences affect different accounts in different ways, and they can operate at different speeds. Typically, economists and policymakers focus on only a few of the more important influences, among which are monetary and fiscal policies. Even within this restricted set of policies, the effects on the balance of payments are complicated. The effects of fiscal and monetary policies on a country's balance of payments will depend on its other policies   especially toward exchange rates and capital flows   and on the size and wealth of the country and the extent to which it trades with others.

See also aid, international; Bretton Woods system; convertibility; foreign direct investment (FDI); gold standard, international; International Monetary Fund (IMF); international reserves; mercantilism; money supply; Mundell-Fleming model; remittances; World Trade Organization

FURTHER READING

International Monetary Fund. *Balance of Payments Statistics Yearbook*. Washington, DC: IMF. Besides providing a wealth of statistical information on balance of payments, this volume explains the methods and rationale of modern balance of payments accounting.

Viner, Jacob. 1937. *Studies in the Theory of International Trade*. New York: Harper and Brothers. Historical exegesis and analysis of international trade and finance theory from the mercantilists to John Maynard Keynes.

JACK HIGH

balance sheet approach/effects

National authorities, market analysts, and the International Monetary Fund (IMF) traditionally have assessed the financial health of a country on the basis of *flow* variables, such as annual gross domestic product (GDP), the current account, and fiscal balances. Sudden and disruptive capital account crises in Mexico (1994 5), Southeast Asia (1997 8), Russia (1998), Turkey (2001), and Latin America (2001 2), however, called into question the capacity of such metrics to provide a full picture of an economy's vulnerabilities. Signs of impending trouble in these countries might have been spotted earlier through a more careful look at mismatches between the *stocks* of a country's assets and its liabilities; that is, by looking at imbalances within and between a country's sectoral balance sheets. Additionally, once a capital account crisis has begun, changes in the exchange rate, interest rate, and other asset prices can propagate the crisis through their effects on the relative valuations of assets and liabilities within and between sectors.

The balance sheet approach (BSA) represents a framework for identifying stock-based vulnerabilities and the transmission mechanisms between sectors that can turn these weaknesses into full-blown capital account crises. Knowledge of sectoral balance sheet mismatches can aid policymakers in reducing vulnerabilities and identifying appropriate policy responses once a financial crisis unfolds.

Key Balance Sheet Concepts The BSA is principally concerned with providing a comprehensive assessment of the currency and maturity mismatches in a country's assets and liabilities that can trigger large adjustments in capital flows. Whereas traditional flow-based analyses of an economy have focused on the gradual buildup of unsustainable fiscal and current account positions over a defined period, the BSA looks at imbalances in the stocks of assets and liabilities, such as debt, foreign reserves, and loans outstanding at a certain point in time. Although the two approaches are obviously interrelated, since stocks are the product of both flows and valuation changes, the BSA focuses on how misalignments in stocks can lead to the sudden changes in flows that presage liquidity and even solvency problems.

An application of the BSA begins by looking at a country's consolidated external balance sheet; that is, its position vis-à-vis nonresidents. This consolidated balance sheet summarizes the external debts of a country's public and private sectors relative to their external assets. The consolidated balance sheet's level of aggregation can, however, mask considerable imbalances between and within sectors that could trigger disorderly adjustments. For instance, a country's consolidated balance sheet does not show foreign currency debt between residents, but such debt can trigger an external balance of payments crisis if the country's government needs to draw on its reserves to roll over its domestically held hard currency debt. In fact, one of the key insights of the BSA is that cross-holdings of assets between residents can create internal balance sheet mismatches that leave a country vulnerable to an external balance of payments crisis.

In order to implement the BSA, an economy can be disaggregated into a set of interlinked sectoral balance sheets. The exact disaggregation used should

depend on the issues under analysis and the data available. A basic matrix of interrelated balance sheets would break down assets and liabilities by maturity and currency, and would include the public sector (i.e., central bank, government, and public corporations), the private financial sector (i.e., principally banks), and the nonfinancial private sector (i.e., companies and households), as well as an external (nonresident) sector to which all three of these domestic sectors are linked. Sectoral balance sheets are interrelated in that one sector's liabilities are by definition the assets of another sector (see figure 1).

Real assets, such as plants and materiel, which are often a major element of public assets, are not included in a BSA matrix since they are not sufficiently liquid to be called on in a crisis. Consequently, a BSA-style analysis provides a snapshot of an entity's net financial position and its capacity to draw on liquid assets in response to pressure for repayment of its liabilities. It is not intended as a method for calculating the net worth or solvency of an economy or sector, nor does it serve as a precise early warning system. Rather, the BSA reveals vulnerability to a sudden rebalancing of stocks without necessarily pro-
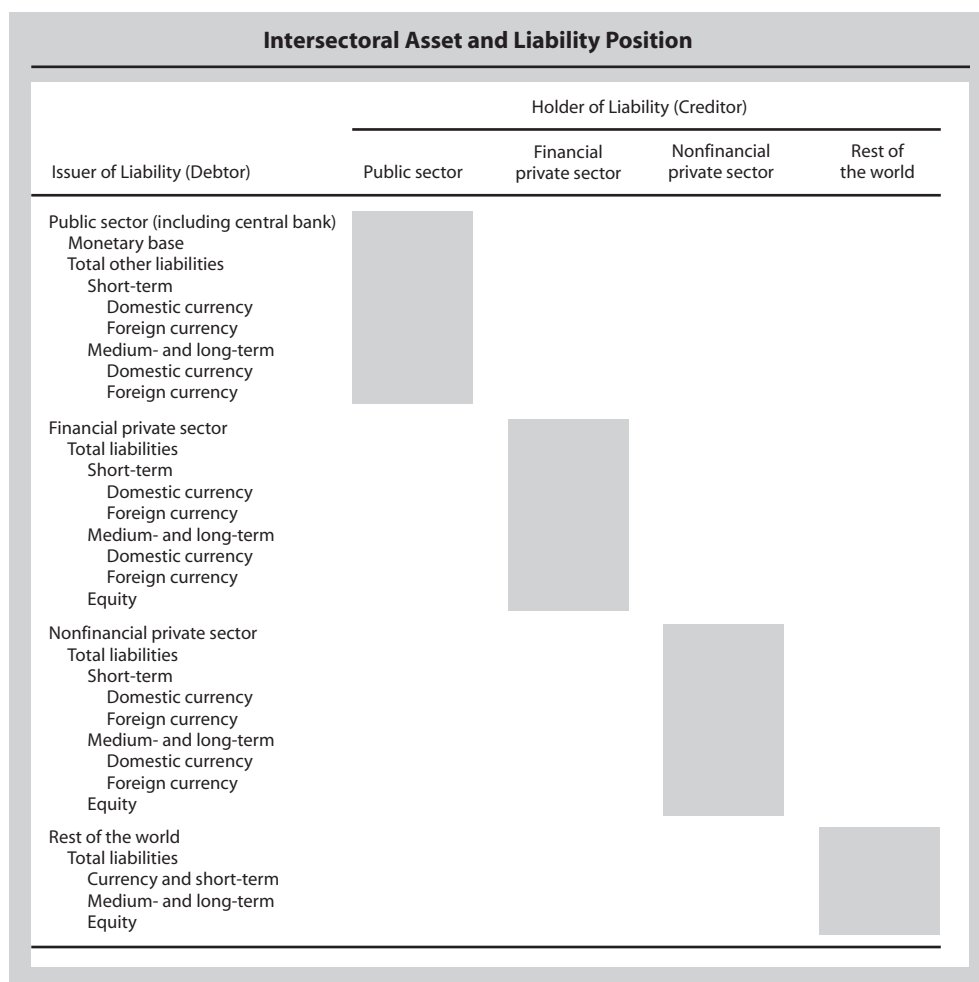
Figure 1
Intersectoral asset and liability position. Source: Rosenberg et al. (2005).

viding an indication of the probability that such an event could occur. In this regard, even partial data can provide useful insights into potential vulnerabilities.

Balance-sheet mismatches do not in and of themselves lead to crises. They simply create the conditions under which a random shock is likely to inflict damage on an economy. Maturity mismatches in Russia in 1998 left its balance sheets vulnerable to rollover and interest rate risks, but commodity price shocks and contagion from the Asian crisis were the actual triggers of the country's crisis. Conversely, Lebanon managed to avoid a financial crisis through 2007 despite years of marked exchange rate, rollover, and interest rate risks connected to gross public debt levels in excess of 180 percent of GDP. Unwavering investor confidence proved an effective shield against real shocks.

The particular vulnerability of emerging markets to sudden capital flow reversals has dictated an early BSA focus on these countries by the IMF and market analysts. The approach is increasingly used, however, to examine specific vulnerabilities, such as asset bubbles and unfunded pension schemes, in both emerging and more mature market economies.

Taxonomy of Balance Sheet Risks At least three major types of mismatch characterize the bulk of balance sheet risks:

1. Maturity mismatches. Mismatches between long-term, illiquid assets and shorter-term liabilities expose a balance sheet to risks related to both debt rollover and changes in interest rates. If liquid assets do not cover maturing debts, an economy or sector may be shut out of capital markets and unable to cover its debt-service liabilities. Similarly, a sharp change in interest rates can dramatically alter the cost of rolling over short-term liabilities, leading to a rapid increase in the cost of servicing debt.

2. Currency mismatches. A currency mismatch most often arises when a borrower's assets are mainly denominated in domestic currency, but its liabilities are denominated in foreign currency, leaving the borrower's balance sheet vulnerable to a depreciation or devaluation of the domestic currency. In many emerging markets, debtors have been motivated to borrow in foreign currency because it is often cheaper than borrowing in local currency. The choice to borrow in foreign currency is sometimes related to deposit dollarization: banks need to match deposits in foreign denominations with loans in the same currency, increasing their incentive to lend to unhedged local borrowers. Following Argentina's crisis and default in 2001 2, many other highly dollarized countries in Latin America also experienced severe crises as a result of the pervasive currency mismatches created by their financial sectors' need to match dollar deposits with dollar loans.

3. Capital structure mismatches. In the context of the BSA, capital structure refers to the balance between debt and equity in an entity's financing. Compared with debt, equity provides a natural buffer during times of balance sheet stress since dividends can be reduced along with earnings, whereas debt payments remain unchanged regardless of circumstances. Capital structure mismatches can arise when a country finances current account deficits through external borrowing (including from official sources, some of which can be difficult to restructure in the event of a crisis) rather than through foreign direct investment or portfolio equity flows.

These mismatches often combine to increase vulnerabilities. For instance, maturity mismatches in foreign currency can create difficulties if market conditions change and domestic borrowers do not have enough liquid foreign currency reserves to cover short-term foreign currency debt. This is what happened in Uruguay in 2002, when domestic banks had difficulties meeting a run on foreign-exchange-denominated deposits. Similarly, financial entities that borrow short-term funds to invest in longer-term debt instruments would suffer from a rise in interest rates brought about, for example, by an exchange-rate defense or cyclical developments, as occurred in Turkey during 2001. Maturity, currency, and capital structure mismatches combined in the Asian crises of 1997 8. Prior to these crises, the Korean government had severely restricted foreign direct investment, and most capital inflows were financed through foreign-currency, shorter-maturity external debt. Similarly, Thailand's tax regime favored corporate debt over equity, which, combined with an implicit nominal exchange rate peg, also led

the Thai nonfinancial private sector to build up foreign-currency denominated debt with nonresident creditors.

When current and expected liabilities exceed assets, a country, sector, or individual entity may face insolvency. Public-sector solvency is often assessed by looking at the ratio of sovereign debt to GDP or to revenue (as a proxy for the government's ability to service its debt). Similarly, a country's overall solvency is usually measured by the ratio of total debt to GDP or to exports. But such measures can indicate vulnerabilities only when combined with other measures of risk exposure and an assessment of related mismatches.

No single debt-to-GDP ratio indicates that a sovereign or balance of payments crisis is imminent. For instance, countries with identical debt-to-GDP ratios, but different currency and maturity mismatches, often face distinct balance sheet weaknesses. The recent literature on debt intolerance draws heavily on the BSA to explain why developing and emerging markets have tended to run into trouble at much lower debt-to-GDP ratios than advanced countries. On the asset side of the public balance sheet, this research has focused on weak revenue bases and poor expenditure control as probable explanations for such low sustainable public debt thresholds. On the liability side, the literature on "original sin" the inability to borrow long term in local currency highlights the vulnerabilities created by the predominance of foreign-currency debt in emerging markets' public borrowing.

Policy Implications The BSA can aid policymakers in preventing and resolving capital account crises. In support of crisis prevention, the approach systematically identifies balance sheet vulnerabilities and highlights sectors in which liquidity buffers may be wearing thin. The Brazilian authorities, for example, were able to head off a recession following the devaluation of the real in 1998 by shifting balance sheet risks away from the corporate and financial sectors, which were exposed to currency risks, and toward the public sector, which was relatively stronger at that time. They did so by issuing foreign-currency-denominated, interest-rate-indexed sovereign debt.

As overnight rates were hiked to defend the real, private domestic holders of these bonds stood to gain. The Brazilian operation also highlights the cost of such risk transfers: although it helped unhedged firms to weather the devaluation of the real, it burdened the public balance sheet for many years to come.

More generally, the BSA has guided policies to reduce vulnerabilities, such as building asset buffers (official reserves), promoting private-sector hedging instruments, strengthening banking supervision, conducting sound liability management operations, and where appropriate instituting flexible exchange rate regimes to reduce incentives for unhedged exposures. Once a country is in a capital account crisis, an awareness of balance sheet mismatches can help the authorities choose an appropriate policy response. The BSA can be useful in comparing, for example, the costs and benefits of letting a currency depreciate or of defending it with changes to interest rates.

Operationalizing and Extending the BSA Ideally, an application of the BSA begins with the compilation of the data needed to complete the basic 4×4 matrix shown in figure 1. Central banks in some OECD countries currently prepare and publish such balance sheet analyses of their economies. In emerging markets, data for the public and private financial sectors are usually easy to obtain, while data for the nonfinancial private sector are often harder to pin down. When national statistics are lacking, information from the Bank for International Settlements and the IMF on a country's international investment position can sometimes help in compiling the external position and deriving the rest of the matrix. In the context of its surveillance and program work, IMF staff has so far completed BSA-style analyses on more than 20 country cases. Furthermore, balance-sheet-related concepts underpin the IMF's frameworks for debt sustainability analysis and the Financial Sector Assessment Program, a key tool for identifying vulnerabilities in countries' banking sectors.

The basic BSA matrix can be further augmented by including off-balance-sheet items such as contingent claims and derivatives in the assessment of

vulnerabilities. Additionally, a full assessment of the risks arising from balance sheet mismatches should factor in attempts to identify the likelihood of future shocks.

See also banking crisis; capital mobility; contagion; currency crisis; currency substitution and dollarization; early warning systems; financial crisis; global imbalances; hot money and sudden stops; international liquidity; International Monetary Fund (IMF); International Monetary Fund conditionality; International Monetary Fund surveillance; international reserves; original sin; spillovers

FURTHER READING

Allen, Mark, Christoph Rosenberg, Christian Keller, Brad Setser, and Nouriel Roubini. 2002. "A Balance Sheet Approach to Financial Crisis." IMF Working Paper No. 02/210. Washington, DC: International Monetary Fund. This paper launched a systematic application of the BSA to the assessment of countrywide vulnerabilities by the IMF, underpinned by insights from the second and third generation currency crisis models.

Eichengreen, Barry, and Ricardo Hausmann, eds. 2005. *Other People's Money: Debt Denomination and Financial Instability in Emerging Market Economies.* Chicago: University of Chicago Press. A comprehensive 10 essay survey of the balance sheet problems created by the phenomenon of "original sin" in emerging markets.

Goldstein, Morris, and Philip Turner. 2004. *Controlling Currency Mismatches in Emerging Markets.* Washington, DC: Institute for International Economics. A wide ranging survey of the problems associated with mea suring and controlling currency mismatches in emerging countries that proposes a new measure of aggregate ef fective currency mismatch and argues that domestic policies can make a big difference in limiting currency mismatches.

Mathisen, Johan, and Anthony Pellechio. 2006. "Using the Balance Sheet Approach in Surveillance: Framework, Data Sources, and Data Availability." IMF Working Paper No. 06/100. Washington, DC: International Monetary Fund. A practical "how to" guide to com piling a BSA style analysis of countrywide and sectoral vulnerabilities.

Rosenberg, Christoph, Ioannis Halikias, Brett House, Christian Keller, Jens Nystedt, Alexander Pitt, and Brad Setser. 2005. "Debt Related Vulnerabilities and Fi nancial Crises. An Application of the Balance Sheet Approach to Emerging Market Countries." IMF Oc casional Paper No. 240. Washington, DC: International Monetary Fund. A review of the evolution of balance sheet vulnerabilities in emerging market countries dur ing the decade from the mid 1990s to 2005, and a col lection of applications of the BSA to both emerging market crises and potential crises that were averted.

Roubini, Nouriel, and Brad Setser. 2004. *Bail In or Bail Out? Responding to Financial Crises in Emerging Markets.* Washington, DC: Institute for International Econom ics. An analysis of possible responses to emerging market crises informed and underpinned by the BSA.

MARK ALLEN, BRETT HOUSE, AND CHRISTOPH ROSENBERG

■ Balassa-Samuelson effect

It has become conventional wisdom in economics that richer countries tend to have higher overall costs of living than poorer countries. Typically this is measured in terms of the real exchange rate, which compares the consumer price indexes of two coun tries converted to a common currency using the nominal exchange rate. This empirical observation has been referred to as the Penn effect, after the Penn World Tables data used to measure it, or alternatively as the Balassa-Samuelson effect, after the economists who wrote about the observation and endeavored to explain it. One important implication of this obser vation is that it indicates a systematic deviation from the theory of purchasing power parity, which is a building block in exchange rate theory. It indicates that there is a role for economic fundamentals such as relative income levels in explaining long-run real exchange rate behavior.

Although numerous theories have been proposed over time to explain this systematic relationship be tween the real exchange rate and income levels, by far the most influential is that proposed in 1964 in two separate papers by Bela Balassa and Paul A. Sa muelson. The theory is based on the divergence of productivity levels in a world of traded and

nontraded goods, explaining that rich countries specialize in and produce goods that are characterized by higher productivity and that are easily traded internationally. (Because this basic idea is also found in an earlier book by Roy F. Harrod in 1933, the theory is sometimes referred to as the Harrod-Balassa-Samuelson effect. This entry will follow the convention of referring to the empirical observation and the theoretical explanation jointly as the Balassa-Samuelson effect.)

Many early empirical studies failed to find statistical support for the connection between relative prices and income levels. It was even more difficult to find statistical evidence of a linkage to the underlying causal factors that the Balassa-Samuelson hypothesis said should be at work, such as between exchange rates and relative productivity levels (see Officer 1982). It appears, however, that the strength of the Balassa-Samuelson effect has grown steadily over time. Recent statistical studies of the second half of the 20th century find that for a large sample of countries the relationship between relative national price levels and income levels became positive as well as statistically significant only in the 1960s, thus validating the Balassa-Samuelson hypothesis (for instance, see Bergin, Glick, and Taylor 2006). It may not be a coincidence that Balassa and Samuelson began writing on the subject at this time. Further, the correlation between these two variables appears to have quadrupled over the half-century since then, and it is very strongly significant statistically in current data.

The Theory of Balassa-Samuelson How exactly are positive correlations between national price levels and income levels related to the Balassa-Samuelson effect? There is a specific way of explaining these correlations, based on differences in productivity levels across countries and goods. Here is a simple version of the theory with an intuitive example to follow.

Consider two countries, home and foreign, where foreign variables are denoted with an asterisk (*). Let there be two goods produced in these countries, where one good (T) can be traded internationally, and the other is a nontraded good (N). Traditional,

albeit imprecise, examples of this distinction would be manufactured goods as traded and services as nontraded. For simplicity, suppose these goods are produced competitively in each country, using only labor as an input, with wages W and W^* in each country. Denote the labor productivity in each sector as A_T and A_N at home, and A_T^* and A_N^* in the foreign country.

If one assumes that trade is costless for the traded good, its price will be equalized in the two countries. Conveniently, this also pins down the relative wage levels in the two countries, since $W/A_T = P_T = P_T^* = W^*/A_T^*$. The wage levels, in turn, pin down the nontraded goods prices with $W/A_N = P_N$ and $W^*/A_N^* = P_N^*$. Now construct a simple consumer price index, say, where the share of expenditure on nontraded goods in consumption is constant at the value θ in both countries. Then the relationship between the price levels of the two countries is given by

$$\frac{P}{P^*} = \frac{(P_N)^\theta (P_T)^{1-\theta}}{(P_N^*)^\theta (P_T^*)^{1-\theta}} = \left(\frac{P_N}{P_N^*}\right)^\theta = \left(\frac{A_T/A_N}{A_T^*/A_N^*}\right)^\theta.$$

This equation predicts that a country will have a higher overall price level if it is highly productive in traded goods, relative to its own nontraded goods, and relative to the traded goods of the foreign country. If one country is richer than the other, this higher income level can be due to higher productivity in the nontraded goods, the traded goods, or some combination of the two. The theory says that the larger the role of productivity growth specifically in the traded sector, the more likely it will be that high relative income levels will be associated with high relative price levels. On the other hand, if a country is richer due to higher productivity in the nontraded sector, or high productivity equally over both sectors, then the model will not predict that the rich country will have a higher overall price level.

As an intuitive and commonly invoked example of the Balassa-Samuelson effect at work, suppose that the home country is rich because it is very good at producing a manufactured good like automobiles, but it has no productivity advantage relative to the

foreign country in terms of a nontraded service like haircuts. The high productivity of home workers in the auto industry affords them a high wage. But it also requires that the wage be high for haircuts, or else no worker would be willing to provide this service, preferring instead to work in the auto industry. Given that a haircut requires the same amount of labor time in each country, but the wage rate paid to the haircutter is higher at home, it is clear that the price of haircuts will be higher at home. Since the purchase price of autos is the same across countries due to arbitrage through trade, the higher price of haircuts makes the overall cost of living higher in the home country.

Implications and Assessment of the Theory
The Balassa-Samuelson theory is used regularly by economists and policymakers to interpret a range of applied issues. Note that a straightforward extension of the theory from levels to changes would imply that countries with faster growth rates in the traded sector would have real exchange rates that are appreciating over time. For example, it predicts that China or other rapidly developing countries might expect pressure for their real exchange rates to appreciate as a natural counterpart to their rapid growth in productivity. Similarly, the theory predicts that if new accession countries joining the European Monetary Union experience a period of accelerated growth as they catch up to richer European countries, they likewise should expect pressure for real appreciation. Since a monetary union effectively implies that the exchange rate is fixed, this pressure should be expressed in this case as a higher inflation rate for countries with higher growth rates. The principle remains the same: higher rates of growth are associated with a rise in the relative cost of living.

The prevalence of the theory behind the Balassa-Samuelson effect in economics owes much to its elegant explanation of the basic price-income relationship. But it has received criticism for the assumptions needed to derive it. There is evidence that productivity gains, especially recently, are not limited to manufactured goods, but that the wealth of relatively rich countries is in part attributable to signifi-

cant productivity gains in many services, such as information technology and retail. Furthermore, it also appears to be true that many services, especially information services, are becoming more tradable due to new telecommunications technologies. As changes in technology and transportation costs lead to significant changes in the volume of trade and even the types of goods and services that are most traded, it is not entirely clear what the future holds for the Balassa-Samuelson effect.

See also equilibrium exchange rate; exchange rate forecasting; purchasing power parity; real exchange rate

FURTHER READING

Balassa, Bela. 1964. "The Purchasing Power Parity Doc trine: A Reappraisal." *Journal of Political Economy* 72 (6): 584 96. One of two original sources for the theory.

Bergin, Paul R., Reuven Glick, and Alan M. Taylor. 2006. "Productivity, Tradability, and the Long run Price Puzzle." *Journal of Monetary Economics* 53 (8): 2041 66. Documents empirically that the fundamental Balassa Samuelson observation is only a modern phenomenon, but that it has strengthened steadily over time.

Officer, Lawrence H. 1982. *Purchasing Power Parity and Exchange Rates: Theory, Evidence, and Relevance.* Greenwich, CT: JAI Press. Documents weakness of the empirical support for the Balassa Samuelson theory in early data.

Samuelson, Paul A. 1964. "Theoretical Notes on Trade Problems." *Review of Economics and Statistics* 46 (2): 145 54. One of two original sources for the theory.

. 1994. "Facets of Balassa Samuelson Thirty Years Later." *Review of International Economics* 2 (3): 201 26. A retrospective commentary on the theory by one of the original authors.

PAUL BERGIN

■ band, basket, and crawl (BBC)

BBC (basket, band, and crawl) constitute the three pillars of an intermediate exchange-rate regime. A currency that uses all three pillars has its central rate (and margins) defined in terms of a basket of currencies rather than a single currency such as the

dollar. It has a wide band around the central rate, perhaps plus or minus 5, 10, or 15 percent, within which the currency floats, but at the edge of which the central bank is obligated to intervene to prevent the market rate going outside the band. Its central rate crawls rather than jumps; when it needs to move, it does so in a series of small, periodic steps rather than occasional abrupt changes. The three component elements were developed by different economists and were driven by different considerations, so they can be analyzed separately even though they tend to appeal to the same group of economists and to be used by the same set of countries. Although in practice there may be a strong complementarity among the three components, it is possible and helpful to consider them separately.

Baskets Before the break-up of the Bretton Woods system in the early 1970s there was no need to consider pegging to a basket of currencies since the major currencies were stable in relation to one another. It became clear that this was not going to continue after the advent of generalized floating in March 1973. Shortly thereafter the International Monetary Fund (IMF) redefined the value of its artificial reserve asset, the special drawing right (SDR), in terms of a basket of the 16 currencies of countries with more than 1 percent of world exports in the base period. When Jordan began to peg to the SDR, it was therefore indirectly pegging to a basket of currencies. This was the first instance of a peg to a basket.

The theoretical basis of the basket peg evolved shortly thereafter, in a number of papers that sought to develop rules for how developing countries should conduct their exchange-rate policies in a world where the major currencies were floating and therefore moving randomly against one another. The first major paper along these lines was that of Stanley Black (1976). He argued that in a context of generalized floating the strategic variable that affected a country's macroeconomic position was not its bilateral exchange rate with any single currency, but its effective (i.e., trade-weighted) exchange rate. According to Black, the weights in the effective exchange rate (EER) should reflect trade, both exports and imports, in goods and services — in other words,

the EER should be a concept broader than just commodity trade but narrower than all transactions that go through the foreign exchange market, in that it should not include capital flows. He also discussed whether the weights should reflect the currency of denomination or the direction of trade, preferring the latter on the ground that this was more relevant in the longer run. He recognized that there would be institutional costs in pegging to a basket, stemming from the fact that one could not intervene in a basket. (One can argue that the main cost of pegging to a basket accrues not — as Black assumed — because of the enhanced calculation costs of the monetary authority, but because it deprives traders of the possibility under normal circumstances of covering forward by utilizing the forward markets of the country's intervention currency.) If, but only if, a country decided that those institutional costs outweighed the macroeconomic benefits of pegging to a basket, then a country should peg to that single currency which minimizes the variance of its EER.

Several subsequent writers entered the debate. The possibility of a number of developing countries all pegging to the SDR was considered, the advantage being that this would avoid arbitrary changes in their cross-rates. Williamson (1982) argued that there was widespread agreement among the authors surveyed in the article that the peg should be chosen with a view to stabilizing something rather than optimizing anything. Specifically, the peg should be picked so as to minimize the instability in real income and inflation imposed by movements in third currencies that are noise to the domestic economy. This is the philosophy that lies behind the first B of BBC. It is perhaps best exemplified by the policy pursued for many years by the Monetary Authority of Singapore.

In recent years proposals for basket pegs have been advanced primarily in the context of East Asia. Several writers (e.g., Williamson 2005) have argued that the motivation to tie the region's currencies to the dollar arises not out of a mercantilist desire for export surpluses but as a response to the fear of losing competitiveness vis-à-vis one another. The way to overcome this collective action problem while avoiding wasting resources on large current account

surpluses as a by-product of dollar depreciation is for all the countries of the region to use a *common* basket peg. (In the case of floating currencies, countries might treat the common basket as a *numéraire*, the unit in terms of which the value of the currency is defined, though this gives less assurance that the market exchange rate would move in parallel to the basket.)

Band Discussion of a wide band for the exchange rate (then called a widening of the gold points) can be traced back to some of John Maynard Keynes's writings in the interwar period, but in the postwar period it was revived by a Brookings Institution report (Salant et al. 1963) and an influential paper of George Halm (1965). Halm looked to a wide band with permanently fixed central rates to enable exchange rate movements to make a worthwhile contribution to the adjustment of balance of payments positions, while decreasing the need for identical monetary policies in different countries and therefore increasing the ability of monetary policy to achieve internal balance. It was this latter aspect on which most subsequent attention focused.

Later authors have differed sharply from Halm in viewing a wide band as complementary to, rather than competitive with, a crawl of the central rate. Few recent analysts have agreed with him in thinking that variations in the exchange rate within the band could be expected to make a worthwhile contribution to payments adjustment; it is changes in the central rate that were expected to do the heavy lifting in that regard. However, the freedom to vary monetary policy has been emphasized by many writers (McKinnon 1971). Williamson (2000) emphasized that this is one of four reasons for favoring a wide zone. A second reason is the difficulty in identifying the equilibrium exchange rate (a rate that is usually conceived as consistent with a sustainable balance of payments outcome) with any precision; there is no point in distorting macroeconomic policy in order to defend an exchange rate target if there is a chance that the rate is not misaligned in the first place. A third reason for a wide zone is to permit nontraumatic changes in the central rate. The analysis here goes back to Harry Johnson (1970), who showed that there would be no incentive to speculate if the change

in the central rate were sufficiently small that the new and old bands overlapped; the wider the band, the easier it is to satisfy this condition. The fourth reason is the desirability of being able to accommodate strong capital flows in part by changes in the exchange rate rather than compelling them to be met entirely by reserve changes.

In 1988 Paul Krugman showed that a credible band would help the authorities to enlist support from speculators in stabilizing the exchange rate (the definitive version was published as Krugman 1991). The intuitive argument was that as the exchange rate approached the margin, speculators would understand that it was increasingly likely to move back toward the central rate, because if it tried to move beyond the edge of the band this would merely provoke the central bank into intervention rather than lead to a further change in the rate. Speculators would therefore be induced to enter the market and help stabilize the rate. The mere promise of official intervention at the margin would suffice to stabilize the rate, without any need for the central bank to actually intervene. First tests of the experience of the European Exchange Rate Mechanism (ERM) were reassuring: it did indeed seem that expectations within the ERM were mean-reverting, in contrast to those that hold in a floating system. But further tests created doubts: Krugman's model predicted that exchange rates would spend most of their time close to the edges of the band, but this did not seem to have happened in the ERM (Svensson 1992). One could reconcile the findings, for example, by intramarginal intervention, but that cast doubt on whether the target zone was really fulfilling its key purpose of making expectations stabilizing. And the ERM was not altogether credible, as is required for Krugman's theorem to apply.

Crawl A central rate is said to be adjusted according to a crawl if its changes are "small." Just how small is best answered by the Johnson analysis cited earlier: a rate is crawling if the old band (prior to the change in central rate) overlaps with the new band. By that test one would count the pre-1987 ERM central rates as mostly crawling, since, although there was no legal obligation for the bands to overlap, in

practice changes in the central rate were usually small enough to produce such an overlap.

A crawl may have several motivations. Probably the principal aim in practice has been to neutralize differential inflation and prevent a country that is inflating faster than the international norm from eroding its competitiveness. This is what led several Latin American countries (first Chile in 1965, then Colombia in 1967, and Brazil in 1968) to institute a crawl in the 1960s. Nowadays we tend to think that there is not much to be gained by running a high rate of inflation, but countries caught up in high inflation found difficulty in reducing it quickly and were acutely interested in preventing it from undermining their trade performance.

A much less common but surely more constructive purpose has been to neutralize biased productivity growth, for example, the bias in favor of tradables that typically comes as a by-product of rapid productivity growth (the so-called Balassa-Samuelson effect). Chile's crawl allowed 2 percent a year in real appreciation for this effect from 1995 until Chile floated in 1998.

Changes in the central rate may also be motivated by the desire to contribute to balance of payments adjustment. A country that wishes to improve its underlying balance of payments position will usually be advised to seek a more competitive exchange rate as one incentive for adjustment. Unless a government chooses to float its currency, it will have to accept a gradual adjustment, so that the incentive for capital flows can be offset by the interest differential. It was the desire to maintain a reasonable balance in international payments that motivated most of the parity adjustments in the early-phase ERM.

Why a BBC System? Perhaps the most basic argument in favor of the BBC regime is that this is the system best calculated to limit misalignments (defined as deviations of the market exchange rate from its equilibrium value), and that misalignments are the principal drawback of both of the alternative regimes. Fixed rates can become inappropriate through differential inflation, Balassa-Samuelson productivity bias, or a real shock that creates a need for balance of payments adjustment. Fixed rates are appropriate

only where there is reasonable certainty that none of these dangers will materialize: where the economy is small and open so as to satisfy the optimum currency area conditions; where it trades predominantly with the currency area to which it plans to peg; where it is comfortable with the inflation policy of that area; and where it is content to adopt institutional arrangements that will guarantee perpetuation of the fixed rate. Flexible rates follow a random walk: they are frequently pushed away from the level that would support a satisfactory evolution of the real economy. A BBC regime provides guidance as to what is considered the longer-run equilibrium rate and mandates action to bring the rate back to that vicinity when it deviates significantly from it. It is true that one advantage of a float, in comparison to the BBC regime, is that this permits a needed adjustment to be made instantaneously, rather than being strung out over time, with the need to retain an offsetting interest differential. There is no guarantee, however, that the private market will choose to make an adjustment at the right time or in the right direction. It is this fear that unguided markets will lead to misalignments that motivates support for the BBC regime.

See also balance of payments; Balassa-Samuelson effect; Bretton Woods system; effective exchange rate; equilibrium exchange rate; exchange rate regimes; exchange rate volatility; hedging; special drawing rights; speculation

FURTHER READING

Black, Stanley W. 1976. "Exchange Policies for Less Developed Countries in a World of Floating Rates." Princeton Essays in International Finance No. 119. Princeton, NJ: Princeton University.

Collignon, Stefan, Jean Pisani Ferry, and Yung Chul Park. 1999. *Exchange Rate Policies in Emerging Asian Countries*. London: Routledge.

Halm, George N. 1965. "The 'Band' Proposal: The Limits of Permissible Exchange Rate Variations." Princeton Special Papers in International Finance No. 6. Princeton, NJ: Princeton University.

Johnson, Harry G. 1970. "A Technical Note on the Width of the Band Required to Accommodate Parity Changes of Particular Size." In *Approaches to Greater Flexibility of*

Exchange Rates: The Bürgenstock Papers, edited by G. N. Halm. Princeton, NJ: Princeton University Press, 280 82.

Krugman, Paul. 1991. "Target Zones and Exchange Rate Dynamics." *Quarterly Journal of Economics* 106 (3) (August): 669 82.

McKinnon, Ronald I. 1971. "Monetary Theory and Con trolled Flexibility in the Foreign Exchanges." Princeton Essays in International Finance No. 84. Princeton, NJ: Princeton University.

Salant, Walter S., Emile Despres, Lawrence B. Krause, Alice M. Rivlin, William A. Salant, and Lorie Tarshis. 1963. *The United States Balance of Payments in 1968*. Wash ington, DC: Brookings Institution.

Svensson, Lars E. O. 1992. "An Interpretation of Recent Research on Exchange Rate Target Zones." *Journal of Economic Perspectives* 6 (4) (fall): 119 44.

Williamson, John. 1982. "A Survey of the Emergent Lit erature on the Optimal Peg." *Journal of Development Economics* 11: 39 61.

———. 2000. *Exchange Rate Regimes for Emerging Markets: Reviving the Intermediate Option*. Washington, DC: Institute for International Economics.

———. 2005. "A Currency Basket for East Asia, Not Just China." Policy Briefs in International Economics PB05 1. Washington, DC: Institute for International Eco nomics.

JOHN WILLIAMSON

■ **Bank for International Settlements (BIS)**
The Bank for International Settlements (BIS) was established in 1929, when representatives of the World War I reparations conference set up a com mittee of experts to provide a definitive financial framework for German war reparations. The BIS was chartered the following year in Switzerland under an international convention. The new institution ap pealed to different constituencies with different in terests. The Germans wanted to increase their exports and link the size of their obligations to capacity to pay; the French sought to replace a debt of the German government to private investors with a debt owed directly to the French government and, ulti mately, to thwart German industrial development; the British were keen to secure enough payments to settle their debt with the United States; and the United States aimed to separate reparations from war debts. But war reparations were only one part of the BIS mission; the other involved extending and deepening cooperation among central banks.

From the beginning, the BIS was a club of central bankers interested in preserving their independence from finance ministries and governments in general. Emblematic of the distinction between central bankers and governments was the decision of the U.S. government not to allow the Federal Reserve System to join the BIS because membership was believed to conflict with the official U.S. position on reparations. Over time, the distinction between central bankers and governments became less sharp, although it still exists; for example, the chair of the Basel Committee on Banking Supervision must be a central bank governor.

The launching of the new institution suffered from poor timing, with much of the industrialized world sliding into economic depression. German war reparations certainly did not help the international economy. In 1931, the Credit-Anstalt, a large bank based in Vienna, went bankrupt, sparking a banking crisis that spread to Germany, Britain, the United States, and most of the countries on the gold stan dard. The BIS acted as a crisis manager and lent to the central banks of Austria, Hungary, Germany, and Yugoslavia, but the treatment was too feeble for the disease. At the core of the problem was an inability of policymakers to understand that feasible cooperation was inadequate to sustain the combination of a gold standard and high employment. The BIS remained a staunch supporter of the gold standard. This posi tion, in addition to allegations that the institution had been too much under (Nazi) German influence before and during World War II, almost brought the BIS to extinction at the Bretton Woods conference in July 1944; it was saved by the Europeans (Toniolo 2005, chap. 8).

The BIS after World War II After World War II, the surviving BIS was out of step with the prevailing economic paradigm and policy prescriptions. The

institution emphasized budget discipline, sound money, free trade, and international monetary cooperation, and had little sympathy for mechanical applications of the standard Keynesian model. Furthermore, central bankers, the BIS's clientele, had lost much of their luster and prestige because of their support of the gold standard in the interwar period. Inevitably, power shifted to finance ministers, who made central banks subordinate to government. Yet, despite this decline in power and prestige, the BIS remained an important forum for central bankers and the international financial community to discuss and find agreement on critical issues of monetary policy and financial stability.

The terms of reference for cooperation have changed over time. In the interwar period, cooperation meant to sustain and "lubricate" the mechanisms of the gold standard. In the 1950s, the focus was on the mechanism of the European Payments Union, in which member countries offset debits and credits in inconvertible European currencies on the books of the BIS and settled balances in convertible currencies. In the 1960s, the BIS became a central point for the discussion of the tensions that were developing in the gold-dollar exchange standard.

Arrangements were made by central banks and among central banks, sometimes using the BIS as the organizer. Central bankers were searching for ways to defend themselves against speculative attacks. The defense arsenal included mutual guarantees all the way to the prohibition of converting funds placed by foreign speculators. Guarantees were implemented by means of swap agreements among central banks and the stipulation that a central bank would repay borrowed funds at the original exchange rate. The techniques evolved over the years. In the mid-1960s, the BIS began to provide regular reviews and analysis of the Eurodollar and Eurocurrencies markets, as well as quarterly statistics on bank lending in the G10 group of countries (Belgium, Canada, France, Germany, Italy, Japan, the Netherlands, Sweden, the United Kingdom, and the United States) and Switzerland to approximately 100 individual countries.

The demise of the Bretton Woods system in 1973 did not spell the end of international financial crises and the role played by the BIS club. On the contrary, financial liberalization in many nations and the abolition of capital and exchange controls exposed the international financial community to more and bigger crises; large currency, debt, and banking crises occurred in Mexico in 1982 and 1994, Southeast Asia in 1997, Russia in 1998, Brazil in 1999, and Argentina in 2001.

The BIS after Bretton Woods The end of Bretton Woods rejuvenated the BIS through a different agenda: European monetary integration and financial regulation. European monetary integration was the natural extension of the European Payments Union. As to financial regulation, the BIS became involved in it after the failure of a German bank, Bankhaus Herstatt, in 1974. Before this date the fixed exchange-rate regime was buttressed by exchange controls and domestic regulation that reduced the risk of financial crises.

Some central banks are also bank regulators and supervisors (e.g., the Bank of Italy); others share this responsibility with other financial regulatory agencies (e.g., the Federal Reserve System); others retain some responsibilities but the bulk of these lies with other government agencies (e.g., Deutsche Bundesbank); and some have no regulatory role aside from that pertaining to the payments system (e.g., Bank of Canada). In practice, however, it is difficult to separate "narrow" central banking from regulation. Even a "narrow" central bank needs credit information on financial participants to prevent losses to either itself or participants in the payments system. Thus it is not surprising that central bank governors considered financial regulation as part of their mandate (Fratianni and Pattison 2001, 205 6).

The BIS acts as a host and secretariat for various committees on financial regulation, of which the best known is the Basel Committee on Banking Supervision (BCBS), which promulgates international standards and "best practices." The BCBS works with national banking regulators, among others, and also participates more broadly, where required, with

securities regulators and insurance regulators. In 1988, a new regime, the Basel Capital Accord, or Basel I, went into effect: internationally active banks were subject to a somewhat common regime for minimum capital requirements. Basel I linked banks' capital requirements to their credit risk through mandated weights for different categories of bank credit. Basel I was ultimately adopted by more than 100 countries and was deemed a success, despite criticism of the crudeness and political bias of the mandated weights: for example, the preferential treatment given to government debtors and to industrial countries that were members of the Organization for Economic Cooperation and Development (OECD).

For several years now, the financial and prudential regulatory committees meeting in Basel have been working on Basel II, a new international agreement on bank capital measurement and minimum capital requirements (BCBS 2004). The main innovation of Basel II is to assess credit risk as the markets would, in contrast to mandated fixed weights. Market-sensitive credit risk assessment, however, requires expensive investments in sophisticated risk management systems and, in practice, can best be implemented by large and internationally active banks. These systems would pay off in terms of lower capital requirements. Smaller banks, instead, would rely on the cheaper and mechanistic formulas of Basel I. Basel II, which has yet to go into effect, got stuck because of U.S. opposition. Small U.S. banks perceive that large banks will gain a competitive advantage, through lower capital requirements, by the implementation of Basel II and have pushed for changes in it. The proposed changes, in turn, would make investment in sophisticated risk management systems cost ineffective and consequently would frustrate one of the main objectives of the new accord.

In sum, the BIS was created with two initial objectives: handling German war reparations payments and promoting and extending central bank cooperation. The first objective is long gone; the second remains central to the current activities of the BIS. The effectiveness of this institution rests on the confidence and commitments that can be made by a

relatively small number of players who meet frequently and in relative seclusion to ensure a high degree of confidentiality. The geographical expansion of the BIS in Asia and Latin America, in the early part of the 21st century, suggests a strategy of enlarging the size of the club.

See also Bretton Woods system; capital controls; convertibility; currency crisis; euro; European Monetary Union; Federal Reserve Board; financial crisis; gold standard, international; international liquidity; International Monetary Fund (IMF); speculation

FURTHER READING

Basel Committee on Banking Supervision. 2004. *Basel II: International Convergence of Capital Measurement and Capital Standards: a Revised Framework*. Basel: Bank for International Settlements. Summarizes the work of the Basel Committee on Banking Supervision to revise capital standards for internationally active banks.

Fratianni, Michele, and John C. Pattison. 2001. "The Bank for International Settlements: An Assessment of its Role in International Monetary and Financial Policy Coordination." *Open Economies Review* 12 (2): 197 222. Reviews the role of the BIS and its adaptability to the changing international financial structure, from the gold standard to floating exchange rates.

Toniolo, Gianni, with the assistance of Piet Clement. 2005. *Central Bank Cooperation at the Bank for International Settlements, 1930 1973*. New York: Cambridge University Press. Offers the most comprehensive history of the BIS from inception to the end of the Bretton Woods system.

MICHELE FRATIANNI

■ Bank of Japan

The Bank of Japan was established in 1882, under the direction of the Ministry of Finance, to be the sole issuer of convertible notes in Japan. The first substantial revision of the Bank of Japan Law occurred in 1942, when the central bank was given the broad objective to conduct its operations "solely for the achievement of national aims" (article 2). There was no mention of financial or price stability, and the law

was designed to support Japan's wartime military effort (Cargill, Hutchison, and Ito 1997 and 2000). In response to the triple-digit inflation in Japan immediately after World War II, in part due to monetization of government deficits, a change in the Finance Law was made in 1947 that provided the Bank of Japan a degree of independence from government deficit financing (e.g., prohibiting the bank from underwriting government bonds or making loans to the government except under some conditions).

In 1949 the Bank of Japan Law was again amended, mainly to provide an overall management structure to the bank by establishing a Policy Board consisting of five voting and two nonvoting members. The Policy Board was given primary authority for almost every aspect of monetary and financial policy operations of the Bank of Japan. Although the Ministry of Finance held a nonvoting position on the Policy Board, the Bank of Japan Law gave the ministry overall control over the bank and the ability to influence policy decisions, if not to determine them outright. For example, the cabinet had the authority to dismiss the governor and vice governor, and the minister of finance could dismiss executive directors, auditors, and advisors of the Bank of Japan "whenever it is deemed particularly necessary for the attainment of the objective of the Bank" (article 47).

New Bank of Japan Law The basic institutional and legal framework governing the Bank of Japan was unchanged during most of the postwar period until creation of the "new" central bank under the changes of the law in 1998 (Cargill, Hutchison, and Ito 2000). A number of economic and political events prompted the change in the law, including widespread dissatisfaction with the Ministry of Finance's handling of the serious banking problem. In particular, the Ministry of Finance was slow to address the bank's nonperforming loan problem (brought on by the collapse of the bubble economy, recession, and poor lending practices), the liquidation of the *jusen* industry (subsidiaries of financial institutions specializing in real estate loans), and other aspects of its response to the financial crisis that had burdened Japan since the early 1990s.

The 1998 Bank of Japan Law fundamentally changed the formal operating objectives of the central bank, its formal relationship to the government, its role in banking supervision, and other functions. Two elements are central to the reform. First, unlike the 1942 law, the new law specified two operating principles for currency and monetary control: the pursuit of price stability (article 2) and the maintenance of an orderly financial system (article 1). The law clearly states that the Bank of Japan is responsible for price stability and shares responsibility for financial stability with other parts of the government. Second, the bank was given much more autonomy from the government and in particular from the minister of finance. The new law states, for example, that autonomy for monetary and currency control shall be respected, and the Ministry of Finance's power to remove Bank of Japan officers was significantly limited. A number of other important changes were also made in the new law, including in the areas of governance and policy formulation, transparency and accountability, budgeting, lender-of-last-resort functions, exchange rate intervention, and government financing.

Conduct of Monetary Policy The conduct of monetary policy has gone through a number of distinct changes during the postwar period. The years immediately following the war were characterized by a chaotic environment and triple-digit inflation during 1946 48. From the mid-1950s to the early 1970s, Japanese monetary policy operated under the constraints of the Bretton Woods fixed exchange system, a period of very high growth, economic progress, and moderate inflation. The early 1970s were again a chaotic period: buffeted by the first oil-price shock and double-digit inflation (called "wild" inflation), Japan struggled to recover a nominal anchor following the breakdown of the Bretton Woods fixed exchange rate system.

The Bank of Japan then introduced a "money-focused" monetary policy in the mid-1970s that served to gradually lower inflation and then keep it at low levels, combined with very strong economic performance through the 1980s (Hutchison 1986). The Bank of Japan, despite its legal dependence on

the Ministry of Finance, and facing the constraints imposed by the wartime conditions of the 1942 law, achieved an extremely successful policy record from 1975 to 1990. The price stabilization record of the Bank of Japan, combined with sustained real growth of the Japanese economy during this period, attracted international attention. The Bank of Japan appeared to be a stark exception to the conventional wisdom that the legal independence of central banks is necessary to generate good inflation records (Cargill, Hutchison, and Ito 1997). The biggest dilemma the Bank of Japan faced at the end of this period the latter part of the 1980s was how to respond to sharp increases in asset price inflation (in retrospect identifiable as an asset price bubble) without affecting other aspects of the economy, which otherwise seemed to be performing well, with strong economic growth and low price inflation overall.

The 1990s brought a completely new set of challenges for the Bank of Japan. The country faced a period of stagnation from 1991 to 2003, with real gross domestic product growth averaging only 1.5 percent, about half that of the rest of the members of the Organisation for Economic Co-operation and Development and much less than Japan's historical norm. During much of the stagnation Japan experienced prolonged deflation, or essentially zero price changes, and a sustained period of near-zero short-term interest rates when the central bank seemed unable to provide additional stimulus to the economy (Hutchison, Ito, and Westermann 2006). During this latter period, much of it when the Bank of Japan was operating independently under the new law, the bank came under severe criticism for not preventing price deflation and not doing more to stimulate the economy. Whether the Bank of Japan could have, or should have, operated differently during this period remains an open question.

See also Bretton Woods system; European Central Bank; exchange rate regimes; Federal Reserve Board; monetary policy rules; money supply

FURTHER READING

Cargill, Thomas, Michael Hutchison, and Takatoshi Ito. 1997. *The Political Economy of Japanese Monetary Policy.* Cambridge, MA: MIT Press. This detailed account of Japanese monetary and financial history from the post war period through the mid 1990s also provides a his torical overview of central banking in Japan, including the creation of the Bank of Japan in 1882 and the 1942 revision in the Bank of Japan Law.

———. 2000. *Financial Policy and Central Banking in Ja pan.* Cambridge, MA: MIT Press. A detailed history of the Bank of Japan, explaining how the change in the Bank of Japan Law in 1998 came about and why it represents the most fundamental change in the bank since its original charter. All of the quotations from various versions of the Bank of Japan Law come from this volume (chapter 4, titled "The New Bank of Japan"), where the specific sources translated into English are provided.

Hutchison, Michael. 1986. "Japan's 'Money Focused' Monetary Policy." (Federal Reserve Bank of San Fran cisco) *Economic Review* 3 (summer): 33 46. Explains the importance and economic effects of the Bank of Japan's new monetary policy implemented in the mid 1970s.

Hutchison, Michael, Takatoshi Ito, and Frank Wester mann. 2006. "The Great Japanese Stagnation: Lessons for Industrial Countries." In *Japan's Great Stagflation: Financial and Monetary Policy Lessons for Advanced Economies*, edited by Michael Hutchison and Frank Westermann. Cambridge, MA: MIT Press. An account of Japanese monetary policy through the Great Stagna tion (1990 2003) and the road to recovery.

MICHAEL HUTCHISON

■ banking crisis

In modern economies, banks play an important role in mediating between borrowers and lenders. Lenders, which are usually households, often do not have profitable investment opportunities on their own and seek high-quality borrowers from whom they can secure a higher rate of return. Households have difficulty finding high-quality borrowers because the financing needs of borrowers are such that the savings of numerous lenders must be aggregated. The coordination costs are simply too high for households to carry out this aggregation. Also, households usually

do not have expertise in evaluating the quality of borrowers, so the lender-borrower relationship can be plagued by adverse selection problems (that is, lenders who lack sufficient information to make wise selections may be apt to select "bad" borrowers). Collecting the information on borrowers necessary for evaluating their quality is costly. Modern banks can aggregate the individual savings of lenders and, by developing expertise in evaluation of borrowers, can mitigate adverse selection problems, thereby smoothing the transfer of funds from households to borrowers.

Banks can play other important roles, such as lengthening the maturity structure of loans by borrowing short-term and lending long-term. Banks do this by taking short-term deposits of say, 1 month, and making loans of say 3 years. Moreover, financial markets such as stock and equity markets are volatile and expose households that invest their savings in them to market risk. By accumulating reserves, which are not typically linked to the stock market, banks can lessen the market risk for households that instead deposit their savings in banks.

There are times, however, when the banks' ability to serve this intermediary role is curtailed and banks actually raise the risk level in the financial system. Because banks have a mismatch between short-term liabilities (e.g., demand deposits) and long-term assets (e.g., loans), they are vulnerable to liquidity crises. Such vulnerabilities can lead to bank runs, as a critical mass of depositors suddenly withdraws funds, leaving the bank without liquid assets to carry on their normal business, which in turn may cause more depositors to panic and withdraw their funds.

In addition, when banks are improperly regulated, they are prone to excessive lending and loan creation, leading to a "bubble" in asset prices. The collapse of the bubble can result in defaults, which can lead to serious dislocation of the economy, as happened to Japan in the 1990s. Excessive loan creation abetted by improper regulation is also the root cause of the emerging market crisis of the late 1990s.

Causes of Banking Crises A banking crisis occurs when many depositors attempt to withdraw their funds all at once. In the past, banking crises happened with some frequency in Europe and the United States, particularly during the Depression in the 1930s. Banking crises occur with some frequency in emerging markets. For example, during the turbulent decade of the 1990s, many emerging markets   including Mexico, Thailand, Indonesia, Korea, Argentina, Russia, and Turkey   faced simultaneous banking and currency crises.

There are two traditional views of banking crises: the "fundamentals" view and the "random shocks" view (Allen and Gale 2000). The fundamentals view is that banking crises are a natural phenomenon of the business cycle. A recession will typically increase loan delinquencies and reduce bank equity, sharply lowering the value of bank assets. As depositors receive information about banking-sector weaknesses, they will withdraw their funds, leading to banking-sector insolvencies. According to this view, bank runs are not random events, but a result of the ups and downs of the business cycle.

The random shocks view is that banking crises are a result of herding or mob psychology. According to this view, bank runs are largely self-fulfilling prophecies (Diamond and Dybvig 1983). If depositors believe that a bank run is about to happen, they will withdraw their funds all at once and a bank run will occur. If depositors believe that bank runs will *not* happen, then they will stay put and a bank run will not occur. A good rendition of the random shocks view is a scene in the movie *It's a Wonderful Life*, in which depositors hear rumors about a failing savings and loan and mob its window. The depositors calm down and stay put when the character played by Jimmy Stewart confidently stresses that the savings and loan is actually doing fine.

In well-supervised jurisdictions, banking regulators promulgate rules and conduct inspections of banks to ensure that they are prudent in their lending, so that when a downturn inevitably comes, banks' balance sheets are not unduly impaired. A bank with an impaired balance sheet has many loans for which the collection of principal and interest is difficult. Thus, the loans will probably have to be written off as

a loss. Should bank balance sheets become impaired, the country's central bank usually serves as the lender of last resort to banks, so that bank runs will not occur. Deposit insurance can serve the same role as "blanket insurance" by the central bank, since the government will "bail out" depositors at a failing bank.

The Late 1990s Banking Crisis in Japan Japan in the late 1990s had a fundamentals-driven banking crisis in which a crash in Japan's equity and real estate markets led to deterioration in bank balance sheets. Dekle and Kletzer (2006) explore the late-1990s crisis in the Japanese banking system and emphasize three key facts about the Japanese financial system at the time: (1) domestic investment was financed primarily by bank loans; (2) the government provided deposit insurance guarantees to the holders of domestic bank deposits; and (3) prudential regulation and enforcement were weak.

Weak prudential regulation in Dekle and Kletzer's model is interpreted as a failure of the government to enforce loan-loss reserve accumulations by banks against nonperforming corporate loans. Banks thus still make dividend payments to their shareholders against the interest collected on their performing and nonperforming loans, when banks should be foreclosing on firms that are in default on their loans. That is, the banks are paying dividends on even the loan repayments they have not yet received. Nonperforming borrowers are kept afloat by further borrowing from banks. Although the public is aware of the mounting nonperforming assets of the banks, they do not withdraw their deposits because of deposit insurance. In effect, deposit insurance allows the banks to transfer resources from the government to their shareholders. Deposit insurance makes the depositors feel safe, so they will keep their deposits even in a failing bank. In the eventuality that a bank fails, the depositors will be bailed out by the government.

If the government fails to intervene by closing banks before a critical date, then the banks' nonperforming loans will exceed the government's ability to borrow. At that point, there is a bank run by depositors, leading to a banking crisis. This appears to be the story of the Japanese banking crisis, from which Japan began to emerge only in 2006.

Twin Crisis in Emerging Markets Another example of a fundamentals-driven banking crisis is the emerging market crisis of the late 1990s. At that time, many emerging market countries were hit not only with a banking crisis but also with a currency crisis, which is characterized by a rapidly depreciating currency, a sharp outflow of foreign exchange, and a slowdown in fact, almost a depression in the domestic economy. The phenomenon of a banking crisis occurring simultaneously with a currency crisis is called a "twin" crisis.

Perhaps the best description of a twin crisis is by Diaz-Alejandro (1985), who discussed the Chilean crisis of the early 1980s. Like many emerging markets, Chile had a nationalized banking system. In the 1970s, Chile's banking system was privatized, even to the extent that authorities repeatedly warned the public that deposits were not guaranteed. In early 1981, however, following the cessation of credit payments by a troubled Chilean sugar company, the central bank bailed out several private banks to stem incipient bank runs. Realizing that the Chilean Central Bank stood ready to protect domestic bank deposits, foreign capital rushed in to take advantage of high Chilean interest rates.

By January 1983 the value of the peso fell as the official Chilean exchange rate rose from 39 pesos per U.S. dollar to 75 pesos. Chilean companies and banks with dollar-denominated debt came under great stress. Nonperforming loans of banks rose from 11 percent of their capital at the end of 1980 to 113 percent by May 1983. Foreign depositors became worried and started to withdraw their funds, rapidly depreciating the Chilean peso. The Chilean Central Bank had no choice but to intervene again, formally guaranteeing all deposits to stem the bank runs and injecting massive liquidity into the banking system to recapitalize the banks.

The Chilean experience illustrates the dangers of capital market liberalization with blanket government deposit guarantees. Such blanket guarantees can lead to moral hazard (in which a party insulated from risk engages in riskier behavior), with too much

capital flowing in, excessive bank lending, a rise in nonperforming loans, bank failures, and finally, a currency crisis.

See also asymmetric information; bail-ins; bailouts; balance sheet approach/effects; Bank of Japan; capital controls; contagion; currency crisis; deposit insurance; financial crisis; financial liberalization; lender of last resort

FURTHER READING

Allen, F., and D. Gale. 2000. *Comparing Financial Systems*. Cambridge, MA: MIT Press. This is the classic graduate level textbook and monograph on the theory of different financial systems.

Dekle, R., and K. Kletzer. 2006. "Deposit Insurance, Regulatory Forbearance, and Economic Growth: Implications for the Japanese Banking Crisis." In *Economic Stagnation in Japan*, edited by M. Hutchison and F. Westermann. Cambridge, MA: MIT Press, 345 415. This is the first general equilibrium model of the Japanese banking crisis.

Diamond, D., and P. Dybvig. 1983. "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy* 91: 401 19. This is the classic and most referenced paper on bank runs.

Diaz Alejandro, C. 1985. "Goodbye Financial Repression, Hello Financial Crash." *Journal of Development Economics* 19 (1 2): 1 24. Although largely descriptive, this is the first paper on how a banking crisis can lead to a currency crisis.

ROBERT DEKLE

■ Basel Convention

The Basel Convention on the Control of Transboundary Movements of Hazardous Wastes and their Disposal is an international agreement that governs cross-border shipments of toxic waste. Adopted in 1989 and entering into force in 1992, the agreement was a direct response to the growing problem of unregulated waste shipments in the 1980s, particularly between rich and poor countries. Waste exports to developing countries emerged at that time as a result of rising disposal fees in rich countries, the need for foreign exchange in poor countries, and an increasingly fluid global economy that facilitated the trade in waste.

The United Nations Environment Program sponsored negotiations on the Basel Convention in the late 1980s following public outcry in response to waste export scandals involving poor countries. The agreement covers toxic, poisonous, explosive, corrosive, flammable, eco-toxic, and infectious wastes. As of October 2007, there were 170 parties to the Basel Convention.

The objectives of the Basel Convention are to reduce the transboundary movement of hazardous wastes, encourage the treatment and disposal of hazardous wastes as close as possible to their generation, and minimize the generation of hazardous wastes.

The agreement governs the international trade in hazardous waste through a system of informed notification and consent. Exporting states must notify the importing country in writing before they ship hazardous waste and must receive consent in writing from the importing state before the shipment can be made. This system allows states to decide case by case whether they wish to accept a particular shipment. Parties to the agreement are prohibited from exporting hazardous waste to states that have banned its import. They are also prohibited from exporting waste to nonparties unless a bilateral or regional agreement allows such trade, provided it is disposed of in no less environmentally sound a manner than that called for in the Basel Convention. Parties to the agreement are encouraged to export hazardous wastes to other countries only if they themselves lack the capacity to dispose of it in an environmentally sound manner, or if the importing country considers the waste a raw material.

Immediately after the Basel Convention was adopted, a number of developing countries and nongovernmental organizations (NGOs) criticized it for having a weak control mechanism. Critics had wanted the agreement to place an outright ban on the trade in hazardous waste between rich and poor countries, but the adopted agreement only regulated exports of waste between these groups of countries. In addition, some countries continued to export hazardous wastes

to developing countries for the purpose of recycling, thereby evading the Basel Convention's control mechanism because the material was not identified as hazardous. Throughout the early 1990s environmental NGOs documented a significant number of toxic waste exports to developing countries for recycling purposes. Most of these exports were recycled in environmentally unsound ways, or were in fact not recycled at all and were simply disposed of.

Basel Ban In response to these concerns, developing countries and NGOs worked to strengthen the convention in the early 1990s. In 1995, at the second conference of the parties to the agreement, the parties adopted an amendment to the convention that bans outright the export of hazardous waste destined for either disposal or recycling from Annex VII countries (parties that are members of the European Union or the Organisation for Economic Co-operation and Development, and Lichtenstein) to non Annex VII countries. This amendment, which has come to be known as the Basel Ban, required ratification by three-fourths of the parties in order to come into force. Many assumed that this number would be 62, as there were 82 parties to the Basel Convention when the Basel Ban amendment was adopted. Uncertainty emerged, however, over ratification procedures when the number of ratifications approached 62: whether the amendment required ratification by three-fourths of the parties at the time it was adopted, three-fourths of the parties present when it was adopted, or three-fourths of the current number of parties. It is expected that the 2008 Conference of the Parties to the Basel Convention will adopt a decision that will resolve this ambiguity. As of October 2007 there were 63 ratifications of the Basel Ban amendment.

In the late 1990s, the parties to the Basel Convention also adopted a classification system for hazardous wastes that delineates more clearly which wastes the treaty covers, including those destined for recycling operations. This system has clarified the scope of the convention considerably.

Challenges for Enforcement The Basel Convention has curtailed some types of hazardous waste exports to developing countries, but two major challenges have emerged that illustrate the difficulty of full enforcement.

The first of these challenges concerns the export of ships containing toxic materials to developing countries for scrapping. Since the mid-1990s, industrial countries have exported a number of ships to scrap yards in developing countries, most commonly India and Bangladesh, for decommissioning. These ships often contain highly toxic materials such as polychlorinated biphenyls (PCBs) and asbestos. In developing countries, ships are commonly dismantled in dangerous and environmentally unsound conditions. Because the ships are technically in use when exported and often not designated as waste material until on the high seas or until they reach their final destination, many have escaped control under the rules of the Basel Convention. Environmental NGOs such as the Basel Action Network (BAN) and Greenpeace have launched campaigns to halt this practice by exposing the failure of exporting states to enforce the Basel Convention when toxic ships are exported from their ports. These efforts have had some success, such as the 2006 decision of the French government to halt the export of the *Clemenceau*, which had been destined for scrapping in India.

The second challenge has to do with the export of electronic waste (e-waste), including discarded computers, mobile phones, and other electronic equipment. E-waste, which contains numerous toxic components, including heavy metals and polyvinyl chloride (PVC), is an especially fast-growing waste stream in industrial countries. Much of this waste has been exported to developing countries, such as China and Nigeria, without being subjected to the Basel Convention rules on notification and consent, and despite bans in those countries on the import of toxic waste. Often these wastes are exported under the pretense of reuse, when in practice they are recycled or simply land-filled. The conditions under which these wastes are recycled and disposed of in these locations are dangerous and environmentally unsound. Several factors have contributed to the continued export of e-waste to developing countries, including the facts that it can escape control when it is

exported for reuse even if most of it is broken and beyond repair, and that customs officials lack knowledge about the hazardous nature of second-hand electronic components. Environmental groups such as BAN have called for pretesting of secondhand electronics before shipment to determine whether they are suitable for reuse or are simply hazardous wastes that would then be bound by the rules of the Basel Convention.

At the same time that the parties to the Basel Convention are considering ways to address these continuing problems with the transboundary movement of hazardous wastes, they have broadened the emphasis of the treaty for its second decade. Environmentally sound management of hazardous waste, as well as the prevention or minimization of such waste, has received more attention. Much of the work of the Basel Convention Secretariat now focuses on enhancing technical and administrative capacity to minimize waste generation and manage hazardous waste safely, especially in developing countries.

The Basel Convention has served an important role in the global economy by regulating the trade in what many consider to be "bads" as opposed to "goods." The parties to the agreement have taken considerable steps toward banning the export of toxic wastes from rich to poor countries. Although the difficulties of monitoring and enforcement allow certain types of toxic waste exports to continue, without the Basel Convention it is likely that the problem of unwanted hazardous waste exports to poor countries would be far worse.

See also multilateral environmental agreements; trade and the environment

FURTHER READING

Clapp, Jennifer. 2001. *Toxic Exports: The Transfer of Hazardous Wastes from Rich to Poor Countries*. Ithaca, NY: Cornell University Press. Outlines the economic aspects of the problem of hazardous waste trade between industrial and developing countries.

Kanthak, Judit. 1999. *Ships for Scrap: Steel and Toxic Wastes for Asia*. Hamburg: Greenpeace. First major report on the environmental impact of ship scrapping in developing countries.

Krueger, Jonathan. 1999. *International Trade and the Basel Convention*. London: Earthscan. Examines the legal relationship between the Basel Convention and international trade rules under the World Trade Organization.

Kummer, Katharina. 1995. *International Management of Hazardous Wastes: The Basel Convention and Related Legal Rules*. Oxford: Oxford University Press. Provides a comprehensive and authoritative account of the legal aspects of the Basel Convention.

O'Neill, K. 2000. *Waste Trading among Rich Nations: Building a New Theory of Environmental Regulation*. Cambridge, MA: MIT Press. Examines the particular problem of hazardous waste trade among industrialized countries.

Puckett, Jim, et al. 2002. *Exporting Harm: The High Tech Trashing of Asia*. Seattle: Basel Action Network and Silicon Valley Toxics Coalition. First major study to document the environmental and social implications of e waste recycling in Asia.

———. 2005. *The Digital Dump: Exporting Re Use and Abuse to Africa*. Seattle: Basel Action Network. Major study by the leading environmental group working on waste issues revealing the extent of e waste exports in Africa.

JENNIFER CLAPP

■ beggar-thy-neighbor policies

Beggar-thy-neighbor policies are those that seek to increase domestic economic welfare at the expense of other countries' welfare. What might be called the classic case of beggar-thy-neighbor policies occurs when one country devalues its currency in order to boost its domestic output and employment but, by so doing, shifts the output and employment problem onto other countries. This occurred in the 1930s when, faced with a worldwide recession, countries sought to increase their own output and employment by devaluing their currencies, a policy that would boost domestic output by reducing the demand for imports and increasing the demand for exports. This exacerbated the recessions in other countries, however, and invited the

response of devaluations by other countries and countries became locked into a series of competitive devaluations.

The experience of the 1930s played a significant part in the design of the Bretton Woods system. This system, by restricting the flow of capital internationally, provided countries with sufficient policy autonomy to target domestic output and employment levels without having to resort to changing the exchange rate. One of the aims of the Bretton Woods system in introducing fixed exchange rates was therefore to obviate the need for competitive devaluations and remove the temptation of beggar-thy-neighbor policies.

With the move to flexible exchange rate regimes since 1973 after the collapse of the Bretton Woods system, beggar-thy-neighbor policies or the possibility of them have occasionally resurfaced. For instance, following the Asian crisis of 1997, when the currencies of Thailand, Indonesia, Malaysia, and South Korea all plummeted, China came under considerable international pressure not to devalue the renminbi, which would likely have further destabilized the crisis-affected countries and led to a new round of currency depreciations. China maintained the value of the renminbi. Similarly, one of the arguments made for trading rivals adopting dollar or euro pegs is that it removes the possibility of beggar-thy-neighbor exchange rate policies. The creation of the euro provided similar benefits to its European members.

Historically, the classic term *beggar-thy-neighbor* has been associated with countries devaluing their currencies to increase domestic output and employment at the expense of other countries. More recently, however, the term has been used more generically to describe policies pursued by one country (or jurisdiction within one country) to increase its economic welfare at the expense of other countries (or jurisdictions), whether the mechanism is exchange rate policy, tax policy, competition policy, or foreign investment policy. Any policies that may have negative spillovers for other jurisdictions are now often referred to as beggar-thy-neighbor policies (Guha 2006).

The solution to the use of beggar-thy-neighbor policies in the 1930s was found in the international policy coordination instituted under the auspices of the Bretton Woods system. In the post Bretton Woods period, the International Monetary Fund and the World Trade Organization face the challenge of providing of such coordination.

See also Bretton Woods system; currency crisis; dollar standard; euro; exchange rate regimes; financial crisis; International Monetary Fund (IMF); international policy coordination; World Trade Organization

FURTHER READING

Betts, Caroline, and Michael Devereux. 2000. "International Monetary Policy Coordination and Competitive Depreciation: A Reevaluation." *Journal of Money, Credit, and Banking* 32 (3) (November): 722 45. Argues that the case for international monetary policy coordination, and hence the avoidance of competitive depreciations, depends on the price setting behavior of firms.

Bordo, Michael, and Barry Eichengreen, eds. 1993. *A Retrospective on the Bretton Woods System: Lessons for International Monetary Reform*. Chicago: University of Chicago Press. A comprehensive analysis of the origins and workings of the Bretton Woods system.

Guha, Krishna. 2006. "IMF Hits at 'Beggar Thy Neighbour' Policies." *Financial Times* (March 8). A discussion of new forms of beggar thy neighbor policies.

PAUL BOWLES

■ black market premium

Black market premium refers to the amount in excess of the official exchange rate that must be paid to purchase foreign exchange on an illegal ("black") market. A black market premium typically arises in two different cases: when the official exchange rate is not the rate that would prevail in the commercial market; and when a nation prohibits access to foreign exchange for specified purposes. In the first case, excess demand at the official exchange rate leads to the development of a parallel, unofficial ("black") market in foreign exchange to meet the demand, with

the market rate exceeding the official rate. The difference between this rate and the official rate represents the black market premium, and it may exceed the difference between the official rate and the rate that would clear a legal market, because of the risks to sellers from participating in the unofficial ("black") market. In the second case, demand for foreign exchange for the prohibited activity induces the supply of foreign exchange in an illegal ("black") market, provided those seeking it will pay a sufficiently high price. Because of the risk involved in supplying foreign exchange for the illegal activity, sellers will demand a price higher than the official rate. The difference between this rate and the official rate will again represent a black market premium.

A black market premium has arisen most often in the context of fixed exchange rate regimes. The premium typically arises when a country fixes the value of its exchange rate in relation to another currency irrespective of the rate that would prevail in the commercial market. It is akin to the authorities' fixing a price for a commodity at a non-market-clearing level. Figure 1 depicts the situation.

In figure 1, schedule DD reflects demand for foreign exchange, while schedule SS reflects the supply. Under normal circumstances DD will be downward sloping, meaning that demand for foreign

exchange will be greater as the price (in units of domestic currency) declines. Similarly, SS will slope upward, since additional foreign currency will be supplied to the market only as the price (in units of local currency per unit of foreign currency) increases. Provided normal economic conditions prevail, the market can be expected to clear at price P^*, where the supply and demand schedules intersect. At this price, quantity Q^* of foreign exchange will be bought and sold.

When a nation fixes its exchange rate at a non-market-clearing rate, the normal market mechanism is disrupted. At the official exchange rate, P^{OFF}, demand for foreign exchange, Q^{DO}, exceeds the available supply, Q^{SO}. Those wishing to purchase foreign exchange cannot obtain it at the official price in the commercial market. If they seek to obtain foreign exchange from a private source, rather than using the queuing mechanism established by the authorities, they will need to pay more than the official price. The margin will reflect the scarcity value of the foreign exchange, plus a premium to compensate sellers for participating in an illegal ("black") market. This risk can be depicted by a leftward (upward) shift in the supply curve to $S'S'$, making the market-clearing exchange rate, P^B, likely to exceed the clearing rate in a legal market. The difference between the clearing rate in the illegal market, P^B, and the official exchange rate, P^{OFF}, is the black market premium.

In the second case, represented by figure 2, the premium arises because sales and purchases of foreign exchange for the specific activity are illegal. In this case, shadow demand and supply schedules, depicted by the lines DD and SS, occur. The difference between the clearing price in this market, again called P^B, and the official exchange rate will represent a black market premium. The size of this premium may differ from that in the previous case, depending on how the law affects the demand for and supply of foreign exchange for the restricted purpose. If the activity itself is illegal for example, seeking foreign exchange to buy illegal drugs risks for the buyer may drive down the price somewhat relative to the case where only the seller faces risks. If the restriction exists alongside a market-clearing exchange
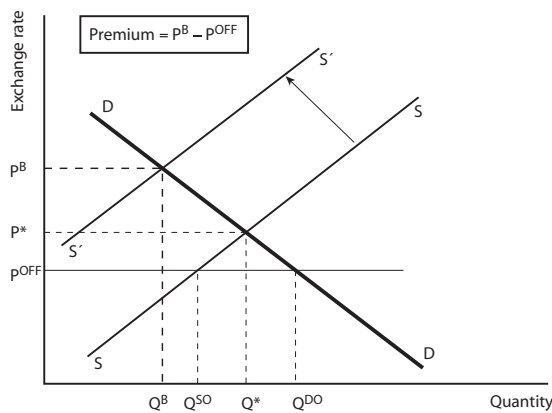
Figure 1
Black market premium when the official rate is not market clearing
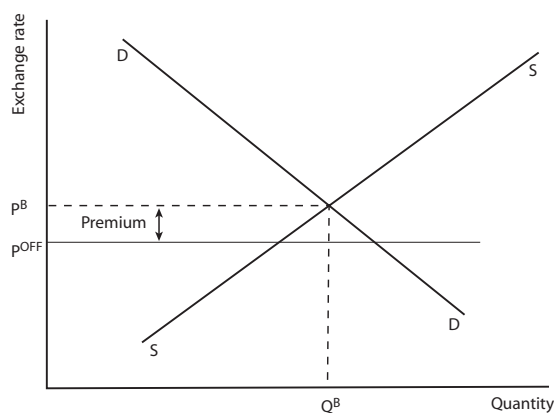
Figure 2

Black market premium for a restricted activity

regime for legal transactions, however, the opportunities for sellers to earn normal profits in the legal market may reduce the supply, and thus raise the price and the size of the black market premium, relative to what would prevail in the previous case of an official rate that is not market-clearing (figure 1).

Whichever situation applies, a black market premium signals a significant imperfection in the exchange market. This, in turn, typically has adverse implications for economic performance. Research has shown that countries with significant black market premiums tend, on average, to have lower rates of economic growth (Fischer 1993). Exchange market imperfections make it hard for foreign exchange to flow to the most productive activities, thus impairing the allocation of resources. In addition, when an overvalued exchange rate is the cause of the problem, firms and individuals must expend additional resources in obtaining foreign exchange. Thus economists typically advocate removing whatever restrictions may be leading to the emergence of black markets in foreign exchange.

See also convertibility; dual exchange rate; exchange rate regimes; foreign exchange intervention; peso problem

FURTHER READING

Fischer, Stanley. 1993. "Role of Macroeconomic Factors in Growth." *Journal of Monetary Economics* 32 (3) (December): 485 512. Argues that larger black market premiums are negatively correlated with economic growth.

Pozo, Susan, and Mark Wheeler. 1999. "Expectations and the Black Market Premium." *Review of International Economics* 7 (2) (May): 245 53. Shows that expectations of devaluation can also affect the size of the premium.

Shachmurove, Yochanan. 1999. "Premium in Black Foreign Exchange Markets: Evidence from Developing Economies." *Journal of Policy Modeling* 21 (1) (January): 1 39. Argues that the size of the black market premium appears to be affected by such factors as the (depreciation adjusted) differential between the country's interest rates and those in major financial markets, and negatively by exports and the level of the official exchange rate.

JOSHUA GREENE

■ Bonn Summit

During the Bretton Woods era (1944 72), the proportion of dollars in world reserves rose steadily. As a result, it was in the countries with a persistent tendency to run balance of payments deficits (the United States and the United Kingdom) that the shortage of world liquidity was observed. In surplus countries (Germany, Italy, and the Netherlands), there was a widespread suspicion that deficit countries were all too ready to accept inflation rather than reduce growth and undergo the discipline needed to restrain their external trade to a position without deficits. In reality, however, world trade growth was not noticeably impeded by reserve shortages when the United States had run more conservative or restrictive policies. But the system was put under significant strain after 1967, when the United States followed policies that greatly increased the supply of dollars and necessitated devaluations. The expansionary policies of the United States, and the inflationary finance of the Vietnam War, are now generally credited with undermining the Bretton Woods system.

In this context, many governments responded to the oil price rise of 1973 by cutting taxes to maintain the level of demand. Governments did this to differing extents, and the industries of different

countries also showed differing abilities to adapt to a new pattern of demand in world trade. This combination of policies led to large current account imbalances and, in particular, a weakness in the dollar. But despite those policies, unemployment rose generally largely because profits bore the brunt of the adverse terms of trade shock, while wages did not adjust. At this point policymakers regarded unemployment as the more serious problem, and the view that potential growth had been permanently reduced was not widely accepted. Meanwhile, inflation, though subsiding from its 1974 peak, remained high in many countries.

Several years of discussion among governments in this climate led to the Bonn Summit agreement of 1978, which epitomized the period. Long discussions and careful staff work finally led to a wide-ranging agreement that embraced, but went beyond, macroeconomic policy. In an effort to offset what was seen as flagging growth, while contributing to better-balanced international payments, Japan and Germany were asked to take stimulative fiscal policy actions. In return, the United States agreed to liberalize its energy prices so that they would rise to world levels, dampen U.S. energy demand, and reduce prices for all oil importers.

This agreement has been extensively criticized. It has been viewed by many as the imposition of American power, pressuring Germany and Japan to accept inflationary policies in order to assist the United States. It was also widely criticized as an example of inappropriate and procyclical fine-tuning of aggregate demand. In fact, the Bonn Summit agreement helped each country to achieve the domestic policies it favored but had been unable to achieve because of political opposition at home (Putnam and Henning 1989). The summit may therefore have been necessary to bring about the policies at the time that they were implemented, but the governing groups in each country favored them anyway. Coordinated policies were not imposed on reluctant governments.

Indeed, given the priorities of the time, the summit measures were not mistaken. Policy preferences did change soon afterward, however. In 1978, the

German government wanted faster growth when faced with forecasts of 3.5 percent real growth and 3.5 percent inflation while unemployment was only 4 percent. Moreover, a strong but temporary surge of inflation, caused by the second oil shock, and the replacement of the governments of the United States and Germany in 1980 and 1982 by Republican and Christian Democratic administrations respectively, led to a substantial reordering of priorities (Holtham 1989).

In practice, the expansionary effects of the fiscal policy changes in Germany and Japan were modest at best. In Germany, they barely offset the automatic claw back through the tax system, and they were accompanied by a tightening of monetary policy. This was acceptable because the summit agreement did not cover monetary policy, although it was clear that any German expansions would have to be bond financed. The fiscal measures therefore had a negligible impact on inflation and a very small effect on the current account. As a result, the main effect of the summit was the deregulation of American oil prices which most commentators now agree was beneficial.

The aftermath of the Bonn agreement looked rather different from the circumstances that brought it about. Unlike the first oil shock, it is implausible to ascribe the second oil shock (1979 80) to monetary expansions or loose policies in the industrial countries. Instead it was triggered by the fall of the shah of Iran. Although this did not lead to a substantial reduction in oil supplies, it caused a surge of speculative buying that drove up the oil price.

Nonetheless, by 1980 popular tolerance of inflation was exhausted, and the threat that a new surge in inflation could again ratchet up the trend inflation rate made for considerable uniformity of view in all countries. This reflected several changes of government, and resulted in a lack of interest in policy coordination plus a perception that each country should concentrate on putting its own house in order.

At this point, there was general agreement among governments on the need to reduce inflation and a much greater readiness to bear the costs in unemployment and recession a readiness that had dem-

ocratic endorsement. In part this readiness was disguised behind an assertion that there would be no lasting output or unemployment costs of eliminating inflation. Indeed, some analysts even suggested that there would be no costs at all so long as the authorities showed enough commitment to their policies, since the private sector would then adjust its expectations to the new regime. There was also some concern, however, that the Bonn Summit's attempt at coordination had failed because governments had failed to carry through their part of the bargain, preferring to free-ride on the stimulus provided by other governments. There is little evidence that this in fact happened. But it is a good example of the potential difficulties in negotiating and then sustaining a program of coordinated policies among sovereign governments.

See also balance of payments; Bretton Woods system; international liquidity; international policy coordination; international reserves; Plaza Accord; Smithsonian Agreement

FURTHER READING

Holtham, G. 1989. "German Macroeconomic Policy and the 1978 Bonn Summit." In *Can Nations Agree: Issues in International Economic Cooperation*, edited by R. N. Cooper, Barry Eichengreen, C. Randall Henning, Ger ald Holtham and Robert Putnam, Washington, DC: Brookings Institution, 141 76.

Putnam, Robert D., and C. Randall Henning. 1989. "The Bonn Summit of 1978: A Case Study in Coordination." In *Can Nations Agree: Issues in International Economic Cooperation*, edited by Cooper et al. Washington, DC: Brookings Institution, 12 140.

ANDREW HUGHES HALLETT

■ brain drain

Brain drain is the emigration of skilled and professional workers (such as engineers, scientists, doctors, nurses, and university professors) from a country. These people emigrate legally, become residents or even citizens of a new country (the host or destination country), and stay there with no intention of returning to the source (sending) country. There are two main legal ways for skilled workers to move to another country permanently: direct immigration through official channels, and indirect migration through overseas education. Direct migration refers to obtaining permanent residency directly because the worker possesses a particular skill or is reuniting with family. Indirect, education migration occurs when a university graduate (or graduate of higher education or a training program that qualifies the worker as skilled) goes to another country to pursue further education, but later chooses to stay in the host country and apply for permanent residency after graduation.

One important feature of the current system of direct and indirect migration (except for family reunion) is that the governments of the host countries usually accept those workers with great skills and talent, because usually only those who are smart and perform well will be able to find new jobs in the host country to support their applications for permanent residency.

Brain drain is usually associated with developing countries, since most of the movement of skilled workers is from developing to developed countries. Many countries in Asia and Africa experience substantial outflow of skilled workers, both in absolute numbers and as a percentage of the initial stock of skilled workers, and this can have an adverse impact on the source countries. Furthermore, the big gaps between the wage rates of skilled workers in developed countries and those in developing countries make brain drain difficult to control.

It is not quite correct to regard brain drain as a problem only for developing countries. Many developed countries also experience outflow of some of their skilled workers. Canada, for example, has suffered a net loss of skilled workers and professionals to the United States, and some people have urged the Canadian government to do something to stem the loss. For developed countries such as Canada, however, while some local skilled workers are moving out, skilled workers from other countries, especially from developing countries, are moving in.

Both government policymakers and economists are interested in the impacts of brain drain on the local economy. A simple analysis of these impacts involves three steps: (1) determining the benefits that brain drain may bring, (2) determining the costs that brain drain may cause, and (3) comparing the benefits and costs of brain drain. This approach applies to both the host country and the source country.

Impacts on the Host Country Brain drain, as a form of permanent immigration, is usually not considered a problem by the host country, and in many cases it is regarded as a positive thing for the local economy. First, it gains workers, which contribute to the gross domestic product (GDP) and gross national product. Second, more skilled workers coming in usually drive down the marginal product of these workers, a phenomenon called the Law of Diminishing Returns. This phenomenon will lead to a drop in the market wage rates of these workers. At any point in time, each additional worker moving in will receive a wage rate equal to the value of his/her marginal product. However, the associated drop in the wage rate will mean that all the workers who moved in before will receive a new wage rate lower than their marginal products measured at the time they moved in. This means that the economy benefits from the gap between what these workers have contributed (their marginal products) and what they receive (the wage rate). Third, skilled workers could generate positive externalities for the society (such as increases in productivity and overall employment), and these immigrants can bring benefits to other parts of the economy not measured by the wage rates they earn.

Externalities (that is, the unintended effects of immigration of skilled workers) take two primary forms. By working with other skilled workers, new immigrants can raise the productivity of local workers through friendly competition and interactions. Also, there is usually some kind of complementarity between skilled workers and unskilled workers so that when a firm hires a skilled worker (such as a doctor or engineer) it also hires a certain number of less skilled workers (such as nurses or technicians) to work with

the skilled worker. This means that the employment of skilled workers could generate demand for unskilled workers. This type of externality is especially helpful in solving the problem of unemployment of unskilled workers.

Brain drain, like other types of permanent migration, involves inflow of people, and thus it can have social and political effects. For example, the government may be concerned about any possible impacts of the inflow of foreign workers on crime, social order, and local education, or about cultural and language differences and conflicts between the new immigrants and local residents. Furthermore, the demographic distribution, the income levels, and the political views of the new immigrants may also have impacts on the local society.

Impacts on the Source Country The impacts of brain drain on the source country could be very different from those on the host country. Consider first the special case in which only one skilled worker moves out so that the local wage rate is not much affected. If before the move the worker is paid his/her value of marginal product, then the impact on *those left behind* (TLB, or the rest of the local economy) will be negligibly small because while the economy loses the contribution of the worker, it saves the payment to the worker. The nonharmful effect of brain drain will no longer hold if (1) the outgoing worker received a wage rate less than what he/she has contributed, or (2) more similar workers move out so that the wage rate rises.

If the outgoing worker received a wage rate less than his/her contribution, the emigration is harmful to TLB because the worker takes with him/her his/her previous contribution, which is bigger than what the rest of the economy is able to save. A worker receives less than his/her (value of) marginal product if there are distortions in the labor market. Two main types of distortion are monopsony (in which a single buyer has disproportionate influence in a market) and externality. A monopsonistic firm will tend to underemploy workers and underpay each worker. The existence of (positive) externality can lead to a wage rate less than the marginal product of labor. As explained earlier, a skilled worker can generate two

types of externality: positive impacts on his/her co-workers through friendly competition and interactions, and complementarity between skilled workers and unskilled workers.

Brain drain can hurt TLB even if there are no distortions (such as monopsony or externality) in the economy. Suppose that a sufficient number of skilled workers move out so that the marginal product of labor increases, and thus the wage rate of the remaining workers increases. The gap between the new, higher marginal product and the original wage implies that TLB have experienced a net loss between the contribution of the emigrating workers and the wages saved by no longer paying them.

The impacts of brain drain on TLB in the source country are not all negative; some of them could be positive. For example, skilled workers working abroad may remit part of their income to their family, relatives, and friends back in the source country. They also serve as examples to younger generations of the benefits of acquiring high-level skills. Some of them may even bring back technologies and investment to the source country.

Brain drain can also have dynamic effects on the source country's economy. First, brain drain means that the country is losing human capital. Since human capital is an important growth factor, brain drain can adversely affect economic growth. Moreover, because skilled workers tend to earn high wages before their departure, they usually have saving rates higher than the average rate in the economy. Thus the outflow of some of these high-income workers could pull down the average saving rate of the remaining population, and this means that the local investment rate and thus economic growth will be hurt.

Emigration of these workers could have other impacts on the rest of the economy. One very important impact is that on education, especially higher education. The prospect of moving to other countries where higher wage rates are offered will increase the incentive of local students to seek higher education and in-depth training. This will in general raise the skill level of workers in the economy, and they could help the economy grow. There are,

however, some problems associated with this increase in the demand for higher education, sometimes referred to as a *brain gain*. First, the demand for higher education could be biased toward technical subjects such as engineering, medicine, science, and computer science, and away from other subjects such as the humanities. Second, there is always an excess demand for education, especially higher education. When there is an increase in the demand for higher education, it is not likely that the government will respond with an increase in the supply. As a result, brain drain may not imply an increase in higher education graduates although it may lead to an improvement in the quality of the graduates because talented children will have bigger incentives to get education.

Brain Drain as a Serious Issue for Developing Countries There are several reasons that the governments of developing countries regard brain drain as a serious problem for their economies. First, the wage gaps between developing countries and developed countries are wide, implying big incentives for the skilled workers in the developing countries to emigrate. Second, the externality associated with skilled workers could be high in these developing countries, mainly because there is usually widespread unemployment of unskilled workers. The complementarity between skilled workers and unskilled workers means that the loss of a skilled worker to another country could lead to a drop in the economy's demand for unskilled workers. Thus large outflow of skilled workers could substantially raise the unemployment rate of unskilled workers. Third, the outflow of different types of skilled workers may not be balanced, and the developing countries could very well be losing more of the types of skilled workers that they need the most   for example, doctors, engineers, and nurses, who can easily apply for immigrant visas to developed countries but also are in very short supply in the source countries. Fourth, most developing countries are in an early stage of development, and skilled workers are needed for the economy to take off. Losing some of these workers could seriously hinder the growth of the economy.

Policy Recommendations In view of the impacts of brain drain on local economies, many economists have suggested policies to protect the welfare of TLB.

Exit Tax One suggestion is that those who want to emigrate would have to pay a tax before leaving the country, and that the tax revenue collected would be distributed to the remaining population in some way. The amount of the tax could be set at a level so that TLB would be adequately compensated, and might include an amount to cover the subsidy on education the government provided to these emigrants. It has been argued in economic theory that the increase in the income of the emigrants should be big enough to cover the compensatory tax. For the emigrants the tax payment could come from their own funds, or loans from relatives, friends, and banks. For example, the former Soviet Union imposed exit taxes on some Jews who immigrated to Israel (although in this case the main reason for the tax may have been political deterrence of emigration rather than economic recovery of investment in the emigrants). China had a policy closer in essence to the exit tax suggested here: it required people who went abroad for education to put down deposits before they were allowed to go, and the deposits would be forfeited if after graduation they did not return to China.

Exit taxation is not common in developing countries, partly because it is difficult to calculate the right tax rate. Another reason is that the amount required to adequately compensate TLB could be big, and people who want to move out may not be able to raise that sum of money. Even though it is argued that the future earnings of the emigrants are big enough to cover whatever is required to compensate TLB, banks will not accept these future earnings to guarantee loans. Thus exit taxes, when implemented fully, could turn out to be prohibitive.

Host Countries to Collect Taxes for the Source Countries Since it is difficult to collect taxes before the emigrants actually earn the new wages, some economists have suggested that the host countries, notably the United States and other developed countries, could impose taxes on the new immigrants. The collected taxes would be returned to the source countries on a regular basis, with the purpose of compensating the remaining population. The merit of this argument is that taxes are collected only after the new immigrants have received income. However, it has a number of practical difficulties.

First, the administration of the tax scheme could be very costly, as a lot of information has to be collected in order to calculate the rates for immigrants from different countries. Second, strictly speaking, the tax scheme requires that people from different countries be subject to different rates. The problem of this requirement is that it could create a lot of administration problems, and it may generate feelings of unfairness among the immigrants. To respond to this point, it has been suggested that the source countries impose only one tax rate on the new immigrants, no matter where they came from. The required rate will be the one that is based on the average amount of compensation needed for all important source countries. The tax revenue collected will then be handed over to some international organization such as the United Nations to spend on programs that promote economic growth and development of developing countries.

Even if the high costs associated with administrating the scheme and the difficulty of implementing the scheme can be overcome, the tax scheme may be against the constitution of most developed countries: it requires that the initial residents in the host countries and new immigrants be subject to two different income tax rates, the higher being on the new immigrants. These developed countries are prohibited from imposing such a two-tier tax system, based on whether people emigrated from another country recently.

Other Policy Recommendations The fundamental way to solve the brain drain problem is to raise the local wage rates, so that skilled workers will have less incentive to emigrate. There are at least three possible ways to raise the local wage rates. First, the government can provide a general wage subsidy to all

skilled workers to close the gap between what they may be able to get elsewhere and what they receive locally. This policy could be more effective in dampening brain drain, but it is costly because substantial subsidies will be needed. That could be beyond the budget capacity of the governments. Furthermore, the main reason why these skilled workers are able to earn higher wage rates is because they have higher marginal products in the host countries because of, for example, the availability of more and better capital and facilities. These can hardly be provided by the source country governments in the near future.

The second way is to substantially promote the growth of the economy so that the income levels of the population are raised. Economic growth of the country will boost not only the GDP and national income of the economy but also the wage rates of the skilled workers. The growth of the economy will substantially diminish the incentive of the skilled workers to move out.

This policy is only part of a long-term solution, as it takes a long time for the economy to have a substantial growth. Furthermore, brain drain has a negative impact on economic growth, as brain drain lowers the stock of human capital in the economy. To promote economic growth while skilled workers are moving out is not an easy task. To do that, the government can invest more in education and/or invest more in infrastructure. The first strategy seeks to slow the rate of depletion of the stock of human capital or even to increase the stock. The second one seeks to promote other growth factors, such as the infrastructure needed to attract foreign direct investment or support high-tech development.

The third policy is to implement a two-tier wage system for skilled workers: one at the prevailing rate in the economy, and a second, higher one for those who choose to come back from abroad, usually after graduation from a training program or higher education. The higher wage scale for the returning skilled workers is to attract them to come back. To carry out this policy successfully, either there must be a sub-stantially large public sector or the government must be willing to use widespread subsidies on the employment of these returning graduates. Many countries such as China have been using this policy to lure students who went abroad for education to come back after graduation.

The use of a two-tier wage system, however, has its own costs. This system requires subsidy payments to these returning workers, and the gap between the two wage scales could alienate those workers who were trained locally and are working with those returning skilled workers.

Another policy to slow the rate of brain drain and to encourage returning immigrants is to attract foreign direct investment. There are three channels through which the domestic demand for skilled workers can be raised. First, attract foreign firms to invest and produce locally. Because foreign firms usually have more advanced technologies than what local firms have, foreign direct investment could substantially raise the demand for skilled workers. Second, recent development of technologies allows local skilled workers (such as telephone operators, computer programmers, and accountants) to be employed by firms located in other countries, without having these workers moving out of the country   a phenomenon called offshoring. Third, local production of products to be exported can increase through a process called fragmentation and outsourcing. As firms in developed countries are able to fragment their production processes, they may move the production of some of the components and intermediate inputs of the final products to developing countries where labor costs are much lower. This will raise the demand for skilled workers in the developing countries.

See also brain gain; brain waste; migration governance; migration, international; remittances

FURTHER READING

Beine, Michel, Frederic Docquier, and Hillel Rapoport. 2001. "Brain Drain and Economic Growth: Theory and Evidence." *Journal of Development Economics* 64 (1): 275 89. A widely cited theoretical and empirical

examination of the possibility of gains due to human capital formation of brain drain.

Bhagwati, Jagdish N. 1979. "International Migration of the Highly Skilled: Economics, Ethics, and Taxes." *Third World Quarterly* 1 (3): 17 30. An early consideration of using taxes as a policy tool to address brain drain.

Bhagwati, Jagdish N., and Koichi Hamada. 1974. "The Brain Drain, International Integration of Markets for Professionals and Unemployment: A Theoretical Ana lysis." *Journal of Development Economics* 1 (1): 19 42. An early theoretical examination of the brain drain phenomenon.

Chang, Shirley L. 1992. "Causes of Brain Drain and Solu tions: The Taiwan Experience." *Studies in Comparative International Development* 27 (1): 27 43. A policy anal ysis of Taiwan's experience of and response to brain drain some decades ago with implications for other de veloping countries.

Martineau, Tim, Karola Decker, and Peter Bundred. 2004. "Brain Drain of Health Professionals: From Rhetoric to Responsible Action." *Health Policy* 70 (1): 1 10. A policy analysis of brain drain in the health professions that draws from case studies of Ghana, South Africa, and the United Kingdom.

Mountford, Andrew. 1997. "Can a Brain Drain Be Good for Growth in the Source Economy?" *Journal of Devel opment Economics* 53 (2): 287 303. A theoretical ex amination of the effect of brain drain on productivity in the source economy.

Wong, Kar yiu, and Chong Kee Yip. 1999. "Education, Economic Growth, and Brain Drain." *Journal of Eco nomic Dynamics and Control* 23 (5 6): 699 726. An "overlapping generations" model of endogenous growth examining the effects of emigration on those left behind with a focus on human capital formation.

KAR YIU WONG

■ brain gain

Brain gain refers to the hypothesis that the emigra tion of advanced students and highly skilled workers may produce domestic incentives for investment in education and skills that offset the human capital losses resulting from the departures. This hypothesis poses a challenge to the more widely held view that such emigration results in a net loss of human capital, or brain drain.

A brain gain may be realized only if several con ditions are met. First, many emigrants must receive substantially higher returns on their education and skills than they would have done had they stayed home. Second, these gains must be perceived by in dividuals in the sending countries who have the in terest and potential to acquire more education and skills, thereby inducing a demand for them. Third, domestic institutions must be able to respond effec tively to this demand. Fourth, some of those who have been induced to acquire more education and skills, and thus add to their human capital stock, by the prospect of emigration, must be unable to emi grate. Finally, this group must be able to put their newly gained education and skills to good use at home.

The first condition is likely to be met quite often. Large international differences in compensation are a driving factor in many migration decisions. Al though not all emigrants are able to put their edu cation and skills to work in the receiving countries, many do. A quarter of all college-educated workers in science and engineering occupations in the United States in 2003, for instance, were born outside the country.

When emigrants experience employment success, reports are likely to filter back to interested com munities in the sending countries, fulfilling the second condition. (Indeed, if these reports are exag gerated, so much the better for the brain gain hy pothesis.) Electronic mail and the World Wide Web have expanded and accelerated the communication channels for such information. A case in point is the close association in the late 1990s and early 2000s between salaries in the U.S. information technology industry and foreign student enrollments at the graduate level in related fields.

The third condition may be the most difficult to fulfill. Effective institutions for education and training tend to be rigid. They require highly skilled teachers who may be difficult to recruit and retain. They may need substantial capital inputs for facilities

and equipment. They may be burdened by red tape and political demands. Even when it is possible to expand such institutions, their quality may decline so significantly that their net contributions to human capital formation are far less than hoped for. New higher education franchises in India, for example, are a far cry from the Indian Institutes of Technology that have shaped high expectations among foreign employers.

If the educational bottleneck is surmounted by a large number of prospective emigrants, it does seem probable that a substantial fraction of them will be frustrated in their attempt to secure a better return on their investment in themselves by seeking skilled work abroad. This condition, the fourth for realizing a brain gain, is enacted through immigration quotas, labor market regulations, and other forms of protection for skilled workers in the receiving countries. Applicants for an employment-based "green card" for permanent residence in the United States, for instance, must typically wait several years for a decision.

The final question is whether these would-be emigrants can put their hard-won knowledge to effective use in the domestic economy. This condition, like the third one, is not necessarily easily met. The profile of the demand for skills, especially highly specialized skills, may be different at home than abroad. Nurses and doctors whose training prepared them to treat middle-class patients in comfortable surroundings provide an illustration; they may choose not to practice their professions rather than serve poor villages where the medical needs are greatest.

The data available for assessing the brain gain hypothesis, although they are improving, are still too poor to allow for anything other than tentative conclusions. The firmest of these conclusions is that a brain gain is more likely in countries with large populations than small ones. Beine et al. (2003), for example, find that China and India benefit from current levels of emigration of college-educated individuals and would benefit even more if outflows were to rise. At the other extreme, Guyana, Haiti, and Jamaica, which have lost more than 80 percent of

their college-educated to emigration, have been made worse off. Although more of the 50 developing countries in Beine et al.'s sample are "losers," the winners contain some 80 percent of the total population.

These tentative findings are intuitively appealing, given the conditions required to realize a brain gain. Large countries seem likely to more easily achieve the economies of scale required to support expansion of and new entry into higher education. They may also benefit more from competition among provinces and localities to provide public education and among private educational institutions as well. National quotas in receiving countries amount to a smaller fraction of the populations of large sending countries than small ones. (Citizens of Jamaica, for instance, are more than a hundred times more likely to be admitted as legal immigrants to the United States than those of India.) Finally, the more diversified economies of large countries may provide more opportunities for "surplus" human capital to be used productively.

Although the conditions for a brain gain may be more likely to be fulfilled in large countries, there are many policies that sending countries of all sizes might undertake to raise the odds of such an outcome. Chief among these are educational policies that provide general skills and economic policies that foster investment and entrepreneurship. Receiving countries, too, have a potential role to play, for instance, by linking their immigration policies to labor market conditions at home and abroad. Neither brain gain nor brain drain is, in the final analysis, an outcome of immutable laws of economics, but rather a construction of policy decision-making within institutional constraints.

See also brain drain; brain waste; migration governance; migration, international

FURTHER READING

Beine, Michel, Frederic Docquier, and Hillel Rapoport. July 2003. "Brain Drain and LDCs' Growth: Winners and Losers." IZA Discussion Paper no. 819. Bonn, Germany: IZA [Institute for the Study of Labor, University of Bonn]. The best available empirical assessment

so far of the brain gain hypothesis at the national level of analysis.

Lucas, Robert E. B. 2004. "International Migration to the High Income Countries: Some Consequences for Eco nomic Development in the Sending Countries." Paper presented at World Bank Conference on Development (April). A balanced and thorough review of the literature on brain gain, brain drain, and related issues.

Stark, Oded. 2004. "Rethinking the Brain Drain." *World Development* 32 (1) (January):15 22. A theoretical ex position of the brain gain hypothesis.

DAVID M. HART

■ brain waste

Brain waste refers to the underuse of migrant knowledge within a host country. In contrast, *brain use* refers to a situation in which migrants obtain jobs commensurate with their educational level and are in a position to make use of their knowledge and skills in the host-country labor market. Greater incidence of brain waste translates into lower levels of brain use and vice versa.

The 1990s were characterized by major geopolitical change and state reconfiguration. These changes put in motion new patterns in the movement of people across boundaries and gave rise to several new trends in international migration (International Organization for Migration 2005). An important new trend is that contemporary migration is concentrated in a few developed countries. In 1970, the developed world received 47 percent of all international migrants whereas 53 percent of migrants moved to developing countries. By 2000, the proportion of international migrants to developed countries had risen to 63 percent. The sharpest increase was in the United States, but Australia, Canada, France, Germany, and the United Kingdom also experienced substantial increases in their migrant populations (International Organization for Migration 2005).

In many of these receiving countries, family reunification has been an important vehicle of admission. In recent years, however, admission of skilled workers has gained prominence. In 2000, more than a half million skilled migrants were selected for admission to the United States, along with more than 100,000 in Australia and the United Kingdom, fewer than 100,000 in Canada, nearly 50,000 in New Zealand, and fewer than 10,000 in France (International Organization for Migration 2005). The growing use of skill level as an admission criterion in destination countries is important to the effective use of labor resources in the world economy.

National Origin and Brain Waste Migrants to Europe and especially the United States tend to be highly educated (Özden 2006). Those who migrate to the United States from Africa, the Middle East, Asia, and Eastern Europe are more educated than their counterparts that set out for Western Europe. In contrast, the United States is by far the destination of preference for Latin American migrants, and their educational levels tend to be lower than the average levels in their home countries (Özden 2006).

A key issue in brain waste research is the extent to which migrants' knowledge and skills are employed in destination countries. Some researchers have indicated that, despite the same level of educational attainment, migrants from different regions show significant differences in job placements in the U.S. labor market. One can discern three categories relating to the incidence of brain waste low, moderate, and high from Özden's (2006) study.

The first category is composed of professional migrants from Asia and Western Europe. It also includes migrants from Japan and Australia. Using 2000 U.S. Census data, Özden (2006) concluded that, in general, migrants from Asia tend to do better in terms of brain use than those from Western Europe. Brain use is lowest (and thus the incidence of brain waste is highest) among migrants from the Republic of Korea (33 percent) and Pakistan (38 percent); comparatively moderate for migrants from the Philippines (40 percent), Taiwan (46 percent), and China (51 percent); and highest among those from India (76 percent). For migrants from Western Europe, Japan, and Australia, brain use is estimated at between 38 percent for migrants from Italy to 69 percent for migrants from Ireland. Migrants from

Japan, the United Kingdom, Germany, Canada, and Australia also do relatively well, with between 59 percent (Japan) and 65 percent (United Kingdom). In this study, brain waste and brain use were measured by assessing the gap between educational attainment and job description in the destination country.

The second category of migrants, characterized by moderate brain waste, is composed of those from various African countries. The incidence of brain use is lowest among professional migrants from Ethiopia (37 percent), and highest among those from South Africa (62 percent). For immigrants from countries such as Nigeria, Ghana, and Kenya, brain waste is in the middle of this range, between 40 percent and 52 percent.

The third category of migrants, characterized by a high degree of brain waste, is composed of professional migrants from Eastern Europe, Latin America, and the Middle East. Although the incidence of brain waste is highest in this category, important differences between regions can be discerned. For migrants from Eastern Europe, brain use is estimated at between 31 percent (former Yugoslavia) and 57 percent (Hungary). For migrants from Latin America, brain use is estimated at between 32 percent (Guatemala) and 51 percent (Brazil). Of all groups, migrants from the Middle East experience the greatest incidence of brain waste, with brain use estimated at between 25 percent (migrants from Morocco) and 46 percent (migrants from Lebanon).

Factors Contributing to Differences in Brain Waste What explains the pattern of brain use among migrants from different regions of the world who possess the same educational credentials? Some brain waste researchers attribute the differences to quality and selection variables (Özden 2006; Matto, Neagu, and Özden 2005). *Quality variables* affect the extent to which an immigrant's human capital is valued within the host country. Two such variables are the source country's expenditure on tertiary education and whether the English language is used in the host country. The greater the expenditure on tertiary education, the better the schools and therefore the better the knowledge and skill level of the migrant. Proficiency in the English language can facilitate

migrant communication in the United States since it is the primary language of business.

Selection variables relate to differences in the abilities of immigrants from different countries. Selection variables predict whether immigrants from a particular country represent the highly educated and therefore professionally competitive, or the minimally educated and therefore professionally uncompetitive, members of their own society. Özden (2006) considers three selection variables (gross domestic product GDP per capita, distance, and conflict) that relate to the source country and one selection variable (migration policy) that relates to the destination country.

The higher the GDP per capita of the source country, the higher the opportunity cost of migration. Therefore, those who might wish to migrate are those who expect to get higher monetary returns on their knowledge and skills overseas. Less-educated individuals from high GDP per capita countries are unlikely to migrate because the costs of migration are greater than the expected returns. In contrast, both highly educated and less-well-educated individuals from low GDP per capita countries may want to migrate to high GDP per capita countries because of the expected economic rewards. Migration from low GDP per capita countries tends to favor the highly educated, however, because they are likely to possess the financial resources required to make a move. The implication of this for differentiating the abilities of migrants from different source countries is that we can expect migrants from high GDP per capita countries to be educated and skilled, while, in general, there is greater uncertainty about the abilities of those from low GDP per capita countries since they represent an education and skill mixture that has to be sorted out in the host country labor market. Further, the process of sorting between skilled and unskilled migrants can be costly for organizations.

Distance between source and destination countries can also have an impact on migrant ability through a self-selection process. Migrants are likely to select a distant destination over a closer one if they believe that they will obtain higher returns on their knowledge and skills in the former. Thus highly

skilled migrants from Africa and the Middle East who expect higher rewards in the U.S. labor market are likely to migrate to the United States rather than to the closer destination of Western Europe. Even in Western Europe, where migration policy has favored family reunification and political asylum seekers, there is a trend toward policies that favor economic migration, making the distance explanation more widely applicable.

Military conflict can have at least two impacts on differences in abilities among migrants. The presence of political instability and military conflict creates an incentive for people with varying educational levels to migrate. This, in turn, contributes to a migrant population with a varying skill and knowledge composition. Military conflicts also affect the allocation of scarce resources from education to military efforts in source countries. This, in turn, diminishes the quality of education and the abilities of migrants from affected countries.

The only selection variable that relates to the destination country is migration policy (Özden 2006). This relates to the openness of immigration policies to migrants from particular countries. The more open the migration policies of destination countries, the more varied the skill composition of migrants. In contrast, migration policies can range from being open to migrants from countries perceived to have highly educated migrants in targeted industries, to being relatively restrictive toward migrants from countries perceived to have a low level of educational attainment among their populations.

Destination Country Labor Markets Although quality and selection variables may play an important role in brain waste, they mask destination-country labor market characteristics that may play an even greater role in explaining the differential patterns of brain waste among migrant professionals. The presumed efficiency of the labor market can hold only if employers in the destination country have perfect information relating to the quality and selection variables pertaining to the source countries of each immigrant. Furthermore, they would also have to use this information rationally to assess the value of identical educational credentials of migrants from different source countries. Research in organizational studies, migration, and cross-cultural adaptation, however, suggest that both structural and cognitive biases on the part of employers and migrants alike play a significant role in shaping perceptions of migrants from different parts of the world and affecting hiring decisions.

Research on diversity in organizations in the U.S. context identifies a wide range of structural barriers faced by nonmajority group members. These include discrimination and prejudice based on a variety of differences (Cox 1993), institutional barriers such as the organizational history in relation to diversity practice, formal hiring practices, and informal organizational processes (Nkomo 1992; Cox 1993). There are reasons to believe that these barriers would also be relevant to other destination countries (de Beijl 1996). Although diversity research has provided significant insight into the impact of these types of barriers on minority members in general, it has generally ignored the unique circumstances of migrants. Research on migration and cross-cultural adjustment suggests that many of the barriers identified in the diversity literature are important in understanding migrant work adjustment. There are other barriers faced by migrants, however, such as language, culture clash, lack of information and knowledge about the labor market, and the inability of potential employers to assess migrant credentials (de Beijl 1996). These barriers affect migrants' ability to adapt and become effectively integrated into the work context as well as the employment and promotion decisions of employers. Most important, these employment decisions on the part of employers fly in the face of any rational calculation regarding the migrants' qualifications. Thus exploring the cognitive and structural biases that play a role in migrant work adjustment would give us greater insight into the factors that may contribute to the differential pattern of brain waste among migrants from different parts of the world identified by Özden (2006).

See also brain drain; brain gain; migration governance; migration, international

FURTHER READING

Cox, T. 1993. *Cultural Diversity in Organizations: Theory, Research, and Practice.* San Francisco, CA: Berrett Koehler. Describes the impact of cultural diversity on organizations, including the social psychological dynamics of prejudice and discrimination in organizations.

de Beijl, R. Z. 1996. "Labor Market Integration and Legislative Measures to Combat Discrimination against Migrant Workers." International Migration Papers No. 8. Geneva, Switzerland: International Labor Office. Describes the various sources of discrimination within host countries that inhibit the ability of migrants to integrate effectively into the labor market.

International Organization for Migration. 2005. "International Migration Data and Statistics." In *World Migration 2005: Costs and Benefits of International Migration,* edited by Population Division of the Department of Economics and Social Affairs, UN World Secretariat. Geneva, Switzerland: IOM, 379 421. Provides a wealth of statistical information concerning trends in international migration in general and for specific regions of the world.

Matto, Aaditya, Ileana Cristina Neagu, and Çağlar Özden. 2005. "Brain Waste? Educated Immigrants in the U.S. Labor Market." World Bank Policy Research Working Paper 3581. Washington, DC: World Bank. Useful empirical exploration of variables that explain brain waste among migrants from different regions and countries in the world.

Nkomo, S. M. 1992. "The Emperor Has No Clothes: Rewriting Race in Organizations." *Academy of Management Review* 17 (3): 487 513. Offers a theory for studying racial dynamics within organizations that takes the unique cultural and historical factors of a setting into account.

Özden, Çağlar. 2006. "Educated Migrants: Is There Brain Waste?" In *International Migration, Remittances and the Brain Drain,* edited by Ç. Özden and Maurice Schiff. Washington, DC: World Bank and Palgrave Macmillan, 227 44. Excellent empirical exploration of migration patterns of highly skilled individuals from select countries and explanation of differences in brain utilization of individuals from select countries.

GELAYE DEBEBE

■ Bretton Woods system

The Bretton Woods agreements negotiated at the United Nations Monetary and Financial Conference held in Bretton Woods, New Hampshire, in 1944 laid the groundwork for the creation of the International Monetary Fund (IMF) and the International Bank for Reconstruction and Development, now known as the World Bank. These institutions provided the framework for the postwar international trade and financial system that came to be known as the Bretton Woods system.

In the early 1940s the Allied powers started to consider the shape of the post World War II international financial and trading system. Influencing their decision on measures to be taken was the assessment of why the trading system had broken down in the 1920s and the causes of the economic depression that had plagued the United States and the United Kingdom in the 1930s. The period had been characterized by record levels of unemployment, excessively onerous war reparations payments imposed in the Treaty of Versailles, and volatile international capital flows, as well as the collapse in the prices of primary commodities.

The response to these problems had been the introduction of policies to protect domestic producers by restricting imports, placing controls on the access to foreign exchange used to purchase imports, introducing preferential tariffs on imports from selected countries, and bilateral clearing and payments arrangements to ensure that imports balanced exports. Other measures included international schemes restricting supplies of primary commodities, as well as national import restrictions, export subsidies, and agricultural price supports. The objective of these policies was to defend domestic income and employment and the stability of domestic financial institutions. Instead, the attempted return to the gold standard was thwarted by sustained financial crises.

Postwar planners recognized the need to coordinate national policy objectives in an interdependent international system and concluded that multilateral institutional structures could eliminate the financial factors that had caused international trade

to be viewed as a threat to national economic welfare. It was generally agreed that a stable international financial and exchange rate system was a prerequisite to successful restoration of free international trade.

The United States and the United Kingdom played the major role in the discussions of the shape of the new system, and both governments contributed proposals of what was to be the first step in this process. The UK proposal, authored by John Maynard Keynes, called for an international clearing union based on an internationally created and managed unit of account. The U.S. proposal, devised by U.S. Treasury official Harry Dexter White, proposed an international stabilization fund, similar to the U.S. Federal Reserve System on an international scale, that would hold reserves of the currencies of all members, from which countries could borrow to support their exchange rates.

The difference in the two proposals reflected the different motivations of the two countries. The United Kingdom was concerned that large payments imbalances under the gold standard had required debtor countries to reverse their balance of payments deficits and protect their declining gold stocks by reductions in domestic income and employment. Surplus countries, on the other hand, were accumulating gold and could easily postpone any action to reduce their surpluses through policies to expand their imports from the deficit countries. The goal of the United Kingdom, as a postwar debtor country, was thus to eliminate this asymmetric adjustment to trade imbalances by ensuring full participation of creditors in the international adjustment process.

The United States, on the other hand, was more concerned by the breakdown in free trade and the manipulation of exchange rates that had occurred in the interwar period as countries attempted to protect themselves from the exchange rate instability caused by large payments imbalances and the instability of international financial flows. The U.S. concern, as the major creditor country, was to prevent deficit countries from resolving their debt positions by using protection and exchange rate manipulation to create trade surpluses.

Article VII of the Atlantic Charter, negotiated by Winston Churchill and Franklin D. Roosevelt at the Atlantic Conference in August 1941 during a secret shipboard meeting, set out the conditions for U.S. support of the British war effort. It reflected U.S. concerns that the postwar trade and financial system be built on a rules-based system of relatively free multilateral trade as opposed to the British approach of trade preferences. The differences were worked out at the Bretton Woods Conference in 1944. The trade and commercial policy aspects of the new system were the subject of discussions at the UN Conference on Trade and Employment held in Havana in 1947.

Creation of the Bretton Woods Institutions
The Bretton Woods Conference created the IMF, largely on the lines proposed in the White plan. On a proposal from the United States, concerned to separate the financing of the postwar reconstruction of Europe from the financing of exchange rate stability, an International Bank for Reconstruction and Development (IBRD) was also created to deal with the longer-term problem of development finance for postwar reconstruction in Europe. The unanimously adopted Havana Charter, proposing the creation of an International Trade Organization (ITO) that would deal with the regulations on commercial policy, create a Commodity Stabilization Fund, and meet the British concerns for symmetric balance of payments adjustments, was never ratified by member states. Instead, a single chapter, the General Agreements on Tariffs and Trade (GATT), which dealt with commercial policy, was the only part of the charter that survived as a series of bilateral treaties to eliminate subsidies and reduce tariffs in support of world trade.

The role of these multilateral institutions with complementary mandates to ensure favorable international trade and financial conditions to support rapid income growth and employment is reflected in their stated objectives. The IMF was created "to facilitate the expansion and balanced growth of international trade, and to contribute thereby to the promotion and maintenance of high levels of employment and real income and to the development of the productive resources of all members as the pri-

mary objectives of economic policy." The ITO charter (Chapter 2, Article II) was even more explicit: "The Members recognize that the avoidance of unemployment or underemployment, through the achievement and maintenance in each country of useful employment opportunities for those able and willing to work and of a large and steadily growing volume of production and effective demand for goods and services, is not of domestic concern alone, but is also a necessary condition for the achievement of . . . the expansion of international trade, and thus for the well being of all other countries." Finally, the role of the IBRD was "to promote the long range balanced growth of international trade and the maintenance of equilibrium in balances of payments by encouraging international investment for the development of the productive resources of members, thereby assisting in raising productivity, the standard of living and conditions of labor in their territories."

This ambitious project of negotiating a new, coherent, and coordinated set of international institutions for the postwar period was never completed, however. The IMF and the IBRD became specialized agencies within the UN system, which allowed them to maintain independent governance mechanisms with representation based on each country's economic weight and financial contribution instead of on the UN General Assembly's rule of one country, one vote. In 1996 the GATT was transformed into the World Trade Organization, an ad hoc international body outside the UN system and without the wider mandate to eliminate discriminatory bilateral agreements as the third pillar of Bretton Woods originally intended for the ITO.

As a result, no multilateral institution was created to deal with the stabilization of primary commodity prices or to oversee trade in agriculture. Instead, official intervention in support of the prices of primary commodities continued in the form of ad hoc agreements dependent on controls over supply, and national and (in the case of European Union countries) regional schemes for the support of agricultural prices and incomes. The most important and most costly examples of the latter were found in industrial countries, and, as in the earlier period, they were generally supported by tariffs and other restrictions on imports, as well as by subsidies and other methods of export promotion. The responsibility for the consequences of fluctuations in developing country export earnings caused by volatility in commodity prices was left to the IMF.

The IBRD, although it was commonly referred to as a "world bank," was not a "bank" because it could not create money, and the majority of the financing for European reconstruction effort was provided by the Marshall Plan. As a result most European countries delayed full observance of their obligations in the international trading system, while the need for finance for reconstruction and trade was met by official aid or through regional multilateral monetary arrangements such as the European Payments Union. The financing problems facing the growing number of newly created democratic developing countries were addressed by the creation of special institutions in the World Bank such as the International Development Association and regional development banks. In addition, the UN Conference on Trade and Development, which had no financing capabilities, provided a forum for the formulation of financing policies designed to benefit developing countries.

Membership in the IMF required a country to maintain a fixed parity for its national currency relative to gold or the dollar. If necessary, countries were to maintain these rates by borrowing reserves temporarily from the IMF and by implementing policies to ensure external equilibrium. Only in the case of a "fundamental disequilibrium," where such adjustment would have meant insupportable domestic hardship, could a country contemplate a change in exchange rates, and such a change was to be negotiated with the other members of the IMF.

Dollar Standard Since most countries did not have the requisite gold stocks after World War II, they chose to fix their exchange rates to the dollar. The United States, with most of the global gold stocks, fixed its parity relative to gold. Nonetheless, the weak external positions of most European countries meant that convertibility at fixed parity for imbalances created by commercial account transactions was not

implemented until the end of the 1950s. By 1960 the Bretton Woods fixed exchange rate system came under pressure when gold traded above its official dollar peg of $35 an ounce on the London gold market. Several years earlier, Yale economist Robert Triffin had suggested that this was the inevitable result of the decision made at Bretton Woods to require members to fix their exchange rates relative to gold or the U.S. dollar. Since all countries except the United States had fixed their currencies relative to the dollar, this created a system that resembled the prewar gold standard that the Bretton Wood system was supposed to replace, as the dollar had simply taken the place of gold.

When U.S. external deficits created a stock of international claims on the dollar that exceeded the U.S. gold stock, the United States could no longer ensure the exchange of dollars for gold at the rate of $35 per ounce. This is precisely what occurred in 1960 when the value of U.S. gold stocks at the official parity of $17.8 billion fell short of the $18.7 billion of outstanding liquid foreign dollar claims. To avoid changing the dollar parity, a series of ad hoc measures to prevent dollar conversion into gold and to increase foreign demand for dollars were introduced during the decade of the 1960s. These included a two-tier gold market separating official and private conversions of currencies into gold; an interest equalization tax on borrowing in the U.S. capital market; Operation Twist, which attempted to produce an increase in short-term rates while keeping long-term rates low in order to support U.S. growth prospects; and new methods of calculating the U.S. balance of payments on the basis of liquidity balances.

In 1966, to provide a supplement to the dwindling U.S. gold stocks and an alternative to the dollar in countries' reserves, the Group of 10 (composed of the ministers of finance of the 10 major industrialized countries) proposed a "special reserve drawing right," which was implemented by the IMF in 1967 with the creation of a facility based on "special drawing rights" (SDRs) in the fund. The SDRs, also known as "paper gold," were simple credit entries in the IMF accounts, distributed to members in proportion to their existing membership quotas in the IMF. They could be used in place of dollars or gold to settle payments imbalances. The value of the SDR was initially defined as the gold equivalent of the dollar (0.888671 grams of fine gold), but after the collapse of the Bretton Woods system the SDR was initially redefined as a basket of 16 currencies. In 1981 it was reduced to 5 currencies, with a revision of the weights and currencies that comprise the basket conducted every five years. Rather than a transaction currency that can be used to finance purchase and sale, it is a unit of account to be used by deficit countries losing foreign exchange reserves in defense of their parity to acquire currencies of other members.

Despite all these attempts to shore up the system, U.S. gold stocks by 1971 had fallen to just over $10 billion against outstanding claims of more than $60 billion. In August of that year the United States suspended convertibility of the dollar to gold at the fixed $35-an-ounce parity and introduced an import surcharge. This represented the de facto confirmation of Triffin's prediction and the end of the Bretton Woods system.

From 1971 to 1974 an ad hoc Committee of the Board of Governors on Reform of the International Monetary System and Related Issues (composed of the 20 executive directors of the IMF and known as the Group of 20) worked on proposals for the resurrection and reform of the Bretton Woods system. This attempt was abandoned, however, after the outbreak of the oil crisis in 1973, and negotiations focused instead on interim arrangements such as a quota increase, the creation of a new governance structure in the form of an "Interim Committee" and a "Development Committee," abolition of an official price of gold, and regulations for the valuation of gold held by the IMF on behalf of its members. The changes were finally completed at a special meeting held in Jamaica in 1976 that eliminated the formal obligation under Article IV for members to fix a parity relative to gold or the dollar. Under the new arrangement each member was bound to "collaborate with the Fund and other members to assure orderly exchange arrangements and to promote a stable system of exchange rates." In particular, each mem-

ber was encouraged to "avoid manipulating exchange rates or the international monetary system in order to prevent effective balance of payments adjustment or to gain an unfair competitive advantage over other members"

Changing Roles of the Bretton Woods Institutions After the reform there was no longer a need for countries to borrow from the IMF in order to support their exchange rates. In theory, under flexible exchange rates countries no longer needed to hold exchange reserves since adjustment to external imbalances would take place through the changes in the relative prices of tradable and nontradable goods produced by exchange rate adjustment. IMF operations were thus concentrated on orderly adjustment by ensuring that flexible currency exchange rates reflected underlying economic forces.

The World Bank's mandate has also changed. As the need for European reconstruction finance disappeared, attention shifted to the financing of large economic infrastructure projects to support developing countries. The bank is an active borrower in private capital markets to finance its projects in developing countries. Dissatisfaction with the development impact of these projects, however, and criticism of widespread corruption in their implementation during the 1980s, led the bank to focus on supporting the creation of markets and institutions in developing countries that would eliminate poverty and support their more active participation in the international trading system.

Thus the two Bretton Woods institutions have lived on beyond their original mandates negotiated in 1944 to become complementary and sometimes conflicting institutions that aim to finance development and eliminate poverty in developing countries.

See also currency crisis; dollar standard; exchange rate regimes; Federal Reserve Board; financial crisis; foreign exchange intervention; gold standard, international; international financial architecture; international liquidity; International Monetary Fund (IMF); international policy coordination; international reserves; special drawing rights; Triffin dilemma; World Bank; World Trade Organization

FURTHER READING

Beyen, Johan W. 1949. *Money in a Maelstrom*. New York: Macmillan. Reflections of a participant in the official international monetary discussions on reform of the international financial system.

Gardner, Richard N. 1956. *Sterling Dollar Diplomacy*. Oxford: Clarendon Press. Classic history of the negotiations between the United States and the United Kingdom that led to the Bretton Woods institutions.

Meier, Gerald M. 1974. *Problems of a World Monetary Order*. New York: Oxford University Press. Presents narrative and important documents involved in the creation and evolution of the Bretton Woods system.

Skidelsky, Robert. 2001. *John Maynard Keynes: Fighting for Freedom, 1937 46*. New York: Viking Penguin. Biography containing an account of Keynes's formulation of his proposal for the postwar international monetary system and his participation in the Bretton Woods Conference.

Tew, Brian. 1970. *International Monetary Co operation, 1945 70*. London: Hutchinson. Historical survey of the various stages of evolution of the international monetary system under fixed exchange rates.

van Dormael, Armand. 1978. *Bretton Woods: Birth of a Monetary System*. New York: Holmes and Meier. Personal account of the discussions and negotiations that produced the Bretton Woods agreements.

Williamson, John. 1977. *The Failure of World Monetary Reform, 1971 74*. Sudbury on Thames, UK: Thomas Nelson. A history and criticisms from a participant in the 1970 Group of 20 discussions on the reform of the Bretton Woods system.

JAN KREGEL

■ bubbles

Bubbles are situations in which asset prices persistently deviate from their fundamental values that is, the prices warranted by the true earning potential of firms. Until recently, economists believed that asset prices as a rule reflected fundamental values, and that bubbles were highly implausible. This view relied on an argument of backward induction: the worth of any asset in the period before the final payoff

date would simply be equal to the discounted value of its face value in the period it expired — that is, terminal value. Extending this argument backward, the value of an asset at any period would be equal to the discounted present value of its future stream of revenue. Otherwise, it was argued, an arbitrage opportunity would arise and the price would be pushed back to its fundamental value by the exploitation of this opportunity by informed traders.

Both the rise of rational expectations theory in economics and the idea that assets prices (such as shares and exchange rate) always incorporate all available public information (efficient market hypothesis in its most widely accepted *semistrong* form) went hand in hand with the backward induction argument, underpinning the view that the best way to understand asset prices changes was by focusing on information about fundamental values.

With the rise of behavioral finance theory since the mid-1980s, however, the view that asset prices equal discounted present value of future streams of revenue has come under critical scrutiny (Shiller 2003). In fact, with the growing skepticism about the efficient market hypothesis, much of the theoretical literature on asset pricing has been in a "vibrant flux" (Hirshleifer 2001). Given this change of paradigm in finance theory, it is hardly surprising that conceptual understanding of bubbles has been changing as well. In fact, in many of the new models with risk-averse traders that have asymmetric expectations it has become next to impossible to define what "fundamental value" is (Allen et al. 1993).

The older, standard view argued that all major asset price variations, including famous historical episodes of bubbles such as the Dutch Tulip Mania (1634–37), the Mississippi Bubble (1719–20), and the South Sea Bubble (1720), were not really "bubbles" in the sense that they were caused by changes in fundamental values (Garber 1990). Although it was always recognized that less than fully rational behavioral traders exist in financial markets, defenders of the standard view have held that informed arbitrageurs would undo any mispricing caused by them even when they did not cancel out each other's effect

(Fama 1970; Malkiel 2003). In this setting, bubbles could arise only in assets that had an infinite horizon (and thus lack a terminal value) because in this instance the backward induction argument did not apply. This case has largely defined the limits of initial theoretical interest in bubbles, going back to the economist Paul Samuelson (1958), and is the foundation of more recent work that now goes under the name of "rational" bubbles (also see the comprehensive survey by Santos and Woodford 1997).

With the rise of behavioral theory, however, the effectiveness of arbitrage became a central issue of contention. Behavioral theorists have produced numerous examples where arbitrage is severely limited in its ability to prevent persistent deviations of asset prices from their fundamental values, bolstering the view that major asset price variations are caused by mispricing as opposed to changes in fundamental values (De Long et al. 1990a, 1990b; Shleifer and Vishny 1997; Allen and Gorton 1993).

Contrary to what is often argued, bubbles are not necessarily the result of irrational trading behavior. A body of more recent work suggests that the main cause of bubbles is the uncertainty about higher-order beliefs, that is, rational agents not knowing what other rational agents will do, where irrationality is not a necessary ingredient (Allen et al. 2006). For instance, in Abreu and Brunnermeier (2003) fully rational traders know that the bubble will eventually burst, but in the meantime generate profits by riding it. They cannot individually bring down the price for lack of sufficient funds, but collectively can if they act in tandem. They all know that a common arbitrage opportunity exists, yet they are uncertain when other traders will begin to act on it. Thus each trader has to determine the right time to exit the market without knowing the exit strategies of other traders, and the varied opinions about the right time to exit cause the bubble to persist. In this setting, some news events can have a disproportionately strong effect beyond what their intrinsic information value would warrant by having the effect of causing traders to synchronize their exit strategies, and thus lead to a precipitous fall in price.

The idea that a bubble bursts only when rational traders' subjective expectations begin to coalesce provides an interesting connection to some of the currency crisis models with self-fulfilling expectations. But the linkages between the new body of work on bubbles by behavioral theorists and international finance so far remain few, despite the rich potential for fruitful connections between the two literatures. Such linkages are made in many economic models (e.g., Allen and Gale 1999; Morris and Shin 1999; Shin 2005), where it is argued that financial liberalization has led to asset price bubbles in numerous countries around the world, and the banking crises that usually accompany currency crises are seen to result from the bursting of the asset price bubble.

See also asymmetric information; banking crisis; carry trade; currency crisis; financial crisis; financial liberalization; interest parity conditions; lender of last resort; speculation

FURTHER READINGS

Abreu, D., and M. Brunnermeier. 2003. "Bubbles and Crashes." *Econometrica* 71: 173 204. Shows that a bubble can exist despite the presence of rational arbitrageurs.

Allen, F., and D. Gale. 1999. "Bubbles, Crises, and Policy." *Oxford Review of Economic Policy* 15 (3): 9 18. Argues that in many recent cases financial liberalization has led to a bubble in asset prices, the bursting of which resulted in banking crises and recession.

Allen, F., and G. Gorton. 1993. "Churning Bubbles." *Review of Economic Studies* 60: 813 36. Argues that agency problems caused by asymmetric information in financial markets cause financial assets to trade at prices different than fundamentals.

Allen, F., S. Morris, and A. Postlewaite. 1993. "Finite Bubbles with Short Sales Constraints and Asymmetric Information." *Journal of Economic Theory* 61: 206 29. Examines the conditions under which asset price bubbles emerge even when all agents are assumed to be rational.

Allen, F., S. Morris, and H. S. Shin. 2006. "Beauty Contests, Bubbles, and Iterated Expectations in Asset Mar kets." *Review of Financial Studies* 19 (2): 719 52. Shows that higher order expectations in a fully rational asset pricing model can give rise to bubbles.

De Long, J. B., A. Schleifer, L. Summers, and R. Wald mann. 1990a. "Noise Trader Risk in Financial Mar kets." *Journal of Political Economy* 98 (4) (August): 703 38. Shows that short trading horizons and risk aversion of informed traders can cause bubbles.

De Long, J. B., A. Schleifer, L. Summers, and R. Wald mann. 1990b. "Positive Feedback Investment Strategies and Destabilizing Rational Speculation." *Journal of Fi nance* 45 (2) (June): 379 95. Shows that rational spec ulation might cause rational traders to act as noise traders, jumping on the bandwagon of an asset price bubble the latter might create.

Fama, E. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance* 25 (1): 283 417. Gives a comprehensive review of theoretical and empirical work on capital markets based on the efficient market hypothesis.

Garber, P. 1990. "Famous First Bubbles." *Journal of Eco nomic Perspectives* 4: 35 53. Argues that the famous bubbles in history can all be explained by changes in fundamentals.

Hirshleifer, D. 2001. "Investor Psychology and Asset Pri cing." *Journal of Finance* 56 (4): 1533 97. Reviews theoretical arguments and capital market evidence bearing on the importance of investor psychology for security prices.

Malkiel, B. 2003. "The Efficient Market Hypothesis and Its Critics." *Journal of Economic Perspectives* 17 (1): 59 82. Defends the efficient market hypothesis against the criticisms leveled at it by the proponents of behavioral theory of finance.

Morris, S., and H. Shin. 1999. "A Theory of the Onset of Currency Attacks." In *Asian Financial Crisis: Causes, Contagion, and Consequences*, edited by P. R. Agenor, M. Miller, D. Vines, and A. Weber. Cambridge, UK: Cambridge University Press, 230 61. Examines the role of uncertainty about other agents' beliefs in currency crises.

Samuelson, P. 1958. "An Exact Consumption Loan Model of Interest with or without the Social Contrivance of Money." *Journal of Political Economy* 66: 467 82. Gives

an early discussion of bubbles that can arise in assets that have an infinite horizon lacking a terminal value.

Santos, M., and M. Woodford. 1997. "Rational Asset Pricing Models." *Econometrica* 65 (1): 19 57. Gives a comprehensive survey of the work based on the notion of "rational" bubbles.

Shiller, R. J. 2003. "From Efficient Markets Theory to Behavioral Finance." *Journal of Economic Perspectives* 17 (1): 83 104. Discusses the evolution of finance theory since the 1980s.

Shin, H. 2005. "Liquidity and Twin Crises." *Economic Notes* 34 (3): 257 77. Explores the role of liquidity in crises where a currency crisis and a banking crisis occur simultaneously and reinforce each other.

Shleifer, A. and Vishny, R. 1997. "The Limits of Arbitrage." *Journal of Finance* 52 (1) (March): 35 55. Discusses how anomalies in financial markets can appear and why arbitrage would fail to eliminate them.

KORKUT A. ERTURK

■ **capital account reversals**
See hot money and sudden stops

■ **capital accumulation in open economies**
Capital accumulation increases the amount of machinery, equipment, and structures available to workers in the economy, thus raising their productivity. Moreover, new capital often embodies technological progress. Hence capital accumulation can be viewed as the most direct way of raising the standard of living.

An open economy enjoys opportunities for capital accumulation that are not available in the absence of international transactions. For one, it does not have to produce all the equipment employed in that country. Importing capital equipment allows a country to take advantage of international specialization in the manufacture of capital goods. For example, East and Southeast Asia produce the bulk of semiconductors and computers, while North America and Europe make most of the airplanes. More generally, a country does not have to specialize in capital goods at all, as trade gives it opportunities to acquire machinery and equipment in exchange for commodities or consumption goods. China is perhaps the most prominent current example of the latter case. Although some researchers have suggested that specialization according to comparative advantage in, say, agriculture, may deprive an economy of the benefits associated with learning by doing, it should be noted that importing capital goods gives a country access to sophisticated technologies, helping it to advance its own technological frontier and learn by imitation.

In addition to the classical static gains from trade, opportunities for trade over time are also important. In a closed economy capital goods have to be manufactured domestically, and resources for their production have to be diverted from other uses, in particular from making consumption goods. This constraint is relaxed in an open economy, where first, capital goods can be purchased from abroad, and second, current consumption does not necessarily have to be sacrificed to make room for investment, since foreign borrowing can partially finance both.

Canonical View Traditionally, the relaxation of the intertemporal budget constraint has been viewed as a great benefit of openness. This is particularly true of emerging markets. A prototypical emerging market country is one where current output per worker is relatively low because there is not enough capital. The diminishing marginal productivity of capital means that if such a country has access to the same technology as industrial countries, the rates of return on investment in the emerging market will be high. The country will be poised for growth, but it faces a dilemma. On the one hand, investment today would raise productivity and therefore the standard of living tomorrow. On the other hand, impatience and intergenerational equity considerations create pressures to bring some of that future prosperity into the present, generating high consumption demand today. Since both consumption and investment goods

have to be produced from current resources, however, it is impossible to manufacture more of both, and hard choices have to be made.

The trade-off is much less stark in an open economy. Either consumption or investment goods or both can be imported. Moreover, the total value of everything that the country produces does not have to equal the value of its absorption (the sum of consumption and investment), as long as the country can finance its trade deficit by borrowing or selling its assets abroad. Of course, debts have to be serviced, and opportunities offered by openness are not unlimited. What happens is that the autarkic requirement that domestic output equal absorption at every point in time is replaced with just one intertemporal constraint that the present discounted value of output equal that of absorption. Hence when new investment opportunities arise and the future looks bright, it is possible to increase both investment and consumption, thus laying the foundation of future prosperity and at the same time enjoying some of its fruits in the present.

Foreign financing of trade deficits can take many forms, all of which, being manifestations of intertemporal trade, involve a gain today for a loss tomorrow. The country's residents can take out foreign loans. They can also sell bonds or shares in their enterprises to foreign portfolio investors. A form of foreign financing particularly important for capital accumulation is foreign direct investment, whereby a foreign investor either builds a new plant in the host country or purchases a substantial enough share in a host country enterprise to participate in its management. In any case, some degree of openness in the capital account is required to realize this benefit, which relies on exchanging not only different goods, but also goods delivered at different points in time and thus can be viewed as a generalization of gains from trade.

Of course, intertemporal trade requires the existence of willing partners. In the canonical view, the counterparts of capital-hungry emerging market countries are advanced economies, where capital-labor ratios are high and, because of diminishing returns, the marginal product of capital may be low. These countries (Japan being one example) may not have enough profitable investment opportunities at home, and they may seek opportunities for investment abroad, particularly as aging residents of these countries save for retirement.

To recapitulate, in the canonical view, when a low-income country has in place conditions for high return on investment and rapid growth and opens up to international trade and capital flows, it will experience an investment boom while also increasing consumption. This is possible because the country is able to finance the resulting trade deficit by borrowing abroad or selling its assets to foreigners. Over time, as its productivity catches up with that in advanced countries and returns to investment decline, its growth decelerates, and trade surpluses replace trade deficits. This description fits broadly the experiences of many countries that experienced rapid growth, including Japan and continental Europe after World War II, South Korea from the mid-1960s through the mid-1980s, and the Central European economies during postcommunism transition.

Dynamics of Growth and Capital Accumulation in an Open Economy Abstracting from technological progress, the rate of growth of output per worker is determined by the pace of "capital deepening," or increases in capital per worker. In a closed economy, that rate depends on available investment opportunities and the residents' preferences regarding the choice between investing and consuming their income. Capital accumulation is gradual, and the standard of living, measured by consumption per capita, converges slowly to a steady-state value determined by available technology. Greater impatience an unwillingness to sacrifice current consumption for the sake of investment implies slower convergence. The presence of technological progress does not alter the essence of this progression, except that the economy converges to a steady-state growth rate rather than a steady-state level of output.

As discussed in the previous section, openness relaxes the budget constraint and could accelerate capital accumulation dramatically. Exactly how much is a contentious question in the open-economy macroeconomic literature. If one extends the classical growth

model the Ramsey-Cass-Koopmans model, where rational, forward-looking households and profit-maximizing firms operate in a perfectly competitive environment to an open-economy setting by assuming free trade and perfect capital mobility, the result will be instantaneous convergence. Perfect capital mobility implies unlimited borrowing at a constant interest rate, so domestic residents will immediately borrow enough money and install enough capital equipment to equalize the domestic rate of return on new investment with that available in the advanced economies they are borrowing from. Assuming they have access to the same technology, output per worker would immediately jump to its steady-state level, equal to that in the most advanced countries. Consumption per capita would permanently remain below the level of advanced countries, reflecting the need to service the debt, but it would also stay constant after a momentary upward transition (or grow at a constant rate in the presence of technological progress).

Needless to say, instantaneous income convergence is not observed in practice. In reality, countries cannot borrow unlimited amounts at a constant interest rate, purchase unlimited amounts of capital equipment, and install it costlessly and instantaneously. Many things get in the way. Borrowing over a certain limit may raise the probability of default or result in too much concentration in the lender's portfolio. Lenders may respond by charging the borrowing country a higher interest rate or by cutting off additional funding. Poor legal systems and the risk of government intervention give rise to doubt about the enforcement of contracts and make lending institutions in advanced economies hesitant to commit overly large sums to emerging market borrowers. Lack of information about local investment opportunities also hinders flows of external finance, and underdeveloped financial systems in emerging markets have a limited capacity to process these flows. In addition, some types of capital, such as buildings, may not be tradable internationally, or at least require some local nontradable inputs, such as construction labor, for their installation. Therefore, if some items in the consumption basket (e.g., services) are also nontradable, the trade-off between

consumption and investment is still relaxed by openness, but less than a model with a single tradable good would imply.

These and other obstacles have been incorporated into open-economy macroeconomic models to moderate the rate of convergence. Details of these frictions and other model assumptions have implications for the evolution of other macroeconomic variables, such as domestic interest rates and the exchange rate. For example, combined with the assumption that production in the tradable sector is more capital intensive than in the nontradable sector, friction in the financial markets such that the interest rate charged on foreign borrowing increases with the amount borrowed in a given period results in gradual convergence of interest rates in emerging and advanced countries and in gradual real exchange rate appreciation in the former, a phenomenon known as the Balassa-Samuelson effect.

Recent Experience The canonical view suggests that capital should flow from advanced countries to emerging markets, where returns are higher. Over the first decade of the 21st century, however, a new pattern has emerged, where trade surpluses in China and some other fast-growing emerging economies are financing trade deficits in a number of advanced economies, most notably the United States. The investment rates are very high in the surplus economies, exceeding 40 percent of gross domestic product (GDP) in China, but national saving is even higher, reflecting the relatively low share of consumption in GDP. While China has attracted a lot of foreign direct investment, it has been purchasing foreign financial assets on a grand scale.

Several explanations have been offered for this reversal of the traditional pattern. Among the factors cited are the relatively weak social safety net in China and some other emerging economies, creating the need for high precautionary private saving; the relatively underdeveloped financial systems in these countries, resulting in the flow of these savings abroad, particularly to the United States, whose innovative financial system has been able to generate attractive assets; and government policy of preventing the exchange rate from appreciating despite the

trade surpluses in order to promote exports and to accumulate international reserves as insurance against external shocks.

The canonical view still holds in many instances. Many advanced economies, for example Germany and Japan, are running current account surpluses, and many emerging market countries, including Central and Eastern European economies and many Latin American economies, are running deficits. High commodity prices, which boost the incomes of many developing countries and reduce their reliance on external finance, may have disrupted the traditional pattern, although this explanation clearly does not pertain to China. On the other hand, the fact that capital does not flow to the poorest developing countries does not really contradict the canonical view, as these countries, for many reasons, do not promise high return on investment despite having little capital.

See also capital flows to developing countries; economic development; foreign direct investment (FDI); growth in open economies, neoclassical models; international income convergence

FURTHER READING

Barro, Robert J., and Xavier Sala i Martin. 2004. *Economic Growth*. 2d ed. Cambridge, MA: MIT Press. An introduction to growth models and empirics.

Brock, Philip L., and Stephen J. Turnovsky. 1994. "The Dependent Economy Model with both Traded and Nontraded Capital Goods." *Review of International Economics* 2 (3): 306 25. A somewhat involved model of growth and capital accumulation in a small open economy.

Klyuev, Vladimir. 2004. *Imperfect Capital Mobility in an Open Economy Model of Capital Accumulation*. IMF Working Paper No. 04/31. Washington, DC: International Monetary Fund. A model that combines capital accumulation dynamics with a financial market friction.

Obstfeld, Maurice, and Kenneth Rogoff. 1996. *Foundations of International Macroeconomics*. Cambridge, MA: MIT Press. A standard open economy macroeconomics text.

Prasad, Eswar, Raghuram Rajan, and Arvind Subramanian. 2006. "Foreign Capital and Economic Growth." Paper presented at a Symposium sponsored by the Federal Reserve Bank of Kansas City. Available at http://www.kc.frb.org/publicat/sympos/2006/pdf/Prasad RajanSubramanian.0811.pdf. An account of the changing pattern of international capital flows and possible explanations.

Rodrik, Dani. 1995. "Getting Interventions Right: How South Korea and Taiwan Grew Rich." *Economic Policy* 10 (20): 53 107. A fascinating account of the period of growth takeoff in East Asia, with a focus on the investment boom.

VLADIMIR KLYUEV

■ capital controls

Capital controls are public policies that aim to curb or redirect flows of financial assets (e.g., bonds, loans, stocks, and foreign direct investments) across international borders, through taxes or various types of quantity restrictions. Governments use these policies as a means to generate fiscal revenues and for other economic and/or political reasons.

From an economic efficiency perspective, an important economic rationale for capital controls usually revolves around a "second-best" argument, that is, a distortion of the operation of free markets that cannot be eliminated. These distortions may be due to asymmetric information (when the two sides of a transaction do not have access to the same information), externalities (unaccounted for consequences of agents' decisions), financial sector problems, or other market pathologies. In such cases when the distortion cannot easily be eliminated, a policy that would have been clearly inferior to free markets (such as capital controls) can be used to counter the cost of the initial distortion.

The first modern era that was to a large extent free of capital controls and in which international financial markets became highly integrated occurred prior to World War I (1880 1914). During the war, and during most of the interwar period that followed, numerous countries, including all of the major participants in the international economy, relied on the heavy use of capital controls. During the post World War II era of 1945 72, capital controls were also

used, based on a set of agreements signed at Bretton Woods in 1944. This usage was supported by the International Monetary Fund (IMF). Under the Bretton Woods system, fixed exchange rates and capital controls protected countries from destabilizing external shocks. The IMF Articles of Agreement allowed countries to retain capital controls, stipulating that countries could not draw on IMF resources to meet a "large or sustained outflow of capital" (IMF Articles of Agreement, Article VI, Section 1a).

The countermovement, which began with the breakdown of the Bretton Woods system in the 1970s, sought to remove government controls and allow markets to operate freely. Most developed countries removed the bulk of their restrictions on capital flows in the 1970s and 1980s, with the United States removing its main capital controls in 1974. A number of countries in Asia moved in the same direction during this period and were followed in turn by several South American economies at the end of the 1980s. The majority of African and Middle Eastern countries did not progress as far in liberalizing their capital accounts at the time. This process gained momentum in the early 1990s, with many developing countries removing most of their control on international capital flows. In the developing economies, this trend was seen as part of the neoliberal "Washington consensus," a set of liberalizing policies that included the decontrol of foreign direct investment.

The Asian financial crisis of 1997–98 fostered a wave of analyses that sought to determine its causes. Many analysts charged that the Asian countries that were hit the hardest had liberalized their capital accounts prematurely under pressure from the IMF. This criticism gave rise to intense skepticism about the wholesale removal of capital controls. This debate has spawned an extensive literature striving to evaluate the economic impact of the various types of capital controls.

Classifying Restraints on Capital Movements Restraints on capital flows may broadly be divided into those that focus on capital account transactions (*capital controls*) and those that focus on foreign currency transactions (*exchange controls*).

Capital Controls These involve constraining one or more elements of the balance-of-payments capital account. In principle they can cover foreign direct investment (FDI), portfolio investment, borrowing and lending by residents and nonresidents, transactions making use of deposit accounts, and other miscellaneous transactions. Within each of these categories, there may be a wide range of possible controls. For example:

FDI by residents abroad or nonresidents domestically can be directly restricted, or restrictions can influence the repatriation of profits and initial capital, and the structure of ownership.

Portfolio investment restrictions can take the form of regulations on the issuance or acquisition of securities by residents overseas or by nonresidents domestically. Limitations on the repatriation of dividends and capital gains and transfers of funds between residents and nonresidents may also exist, as may "market-oriented" tax measures.

Regulations on external debt transactions largely take the form of ceilings or taxes on external debt accumulation by residents and firms (financial and nonfinancial institutions). Special exemptions are often provided in the case of trade-oriented enterprises or on a case-by-case basis, as determined by the regulatory authorities.

Restrictions on deposit accounts may be imposed on foreign currency deposits held locally by residents and nonresidents, or deposits held in local currency by residents abroad or by nonresidents overseas or locally.

Other capital controls entail restrictions on real estate, emigration allowances, and other forms of capital transfers.

Some Key Distinctions Mechanisms for seeking to restrain international capital flows may be applied on a *selective* or *comprehensive* basis; they can be based on *outflows* or *inflows*; they can be either *temporary* or *permanent*; and they can focus on direct *quantitative* controls or on using the *price mechanism* via explicit or implicit taxation.

Selective versus Comprehensive Curbs on capital movements may be more or less extensive. At one end of the range, the capital account could be virtually inconvertible (i.e., comprehensive capital controls). India and China are notable examples in Asia. This being said, it is more typical for a country to impose controls selectively, on one or more items within the capital account. Of the 155 countries surveyed in an IMF study, 119 were reported to have imposed some type of restrictions on certain capital account transactions (Ariyoshi et al. 2000). Of the 119 countries with some controls, 67 were reported to use comprehensive controls. The distinction between selective and comprehensive controls may not be precise, however. For instance, even India and China have relative freedom in some forms of capital movements (such as FDI). The distinction is therefore more of degree than of kind. A generally illiberal regime, that is, one with comprehensive controls, typically has a "positive list" of exceptions to the controls. A generally liberal regime, that is, one that imposes controls selectively, is likely to have a "negative list" of items to be controlled.

Outflows versus Inflows What is the purpose of restraints on capital outflows? First, restraints can slow the speed of capital outflows when a country is faced with the possibility of a sudden and destabilizing withdrawal of capital during a time of uncertainty. Second, they are supposed to break the link between domestic and foreign interest rates, recognizing that a country cannot maintain a flexible exchange rate regime, monetary policy autonomy, and an open capital account all at once (i.e., "impossible trinity"). Thus, crisis-hit economies could conceivably pursue expansionary monetary and credit policies as a means of growing their way out of debt or a recession without having to worry about possible capital flight and the weakening of the currency.

Controls on capital inflows have become more common since the mid-1990s and are meant to minimize the chances of an abrupt and sharp capital reversal (bust) in the future. These are sometimes referred to as "speed bumps" or "sand in the wheels" of the international financial system. Empirical studies have indicated that capital controls have been more effective at preventing "excessive" capital inflows than at stemming capital flight (Mathieson and Rojas-Suarez 1993).

The most prominent example of these controls is the Chilean *encaje*, implemented between 1991 and 1998. The *encaje* was the requirement that a fixed percentage (initially 20 percent) of any short-term capital inflow be deposited in a non-interest-bearing account at the Chilean Central Bank for at least three years. The *encaje*'s aim was to slow down capital inflows, prevent an appreciation of the Chilean peso, and discourage short-term flows and shift more flows into less destabilizing inflows with longer maturities.

Temporary versus Permanent Restraints Temporary restraints are seen as a deterrent to excessive outflows or inflows during an "extraordinary" period. When a country is facing the possibility of capital flight, for example, temporary restraints give policymakers time to make appropriate changes in economic policy. Conversely, temporary restraints may be imposed when an economy experiences unsustainably large capital inflows, due to excessive confidence in the growth prospects of the economy (i.e., "irrational exuberance").

The rationale behind temporary restraints arises from the fear that such capital surges could lead to a loss of competitiveness through a real exchange rate appreciation (sometimes referred to as the "financial Dutch disease phenomenon") (Calvo et al. 1996). In addition, the literature on optimal sequencing of economic liberalization has emphasized the need to reform the financial sector in conjunction with putting in place adequate prudential regulation so as to limit the possibility of systemic risks, before attempting to decontrol capital account transactions. As such, temporary controls may allow reforms to be phased (for instance, see Eichengreen et al. 1999).

Permanent controls are seen as necessary even during "normal" times. The rationale here is that even if all the microeconomic distortions are eliminated and macroeconomic policies are generally sound, certain inherent market failures will cause suboptimal decisions to be made in a decentralized

and free market economy. Insofar as these market failures are prevalent in a laissez-faire economy, they may provide a rationale for permanent, rather than event-specific or transitory, capital restraints.

Direct/Administrative versus Market/Price-Based Restraints Restraints can either directly control market movements or they can be a market-based mechanism that alters the structure of price incentives market participants face, thereby inducing them to modify their behavior. Direct controls can generate such problems as bribery and corruption, high enforcement costs, and the inevitable creation of a black market. These drawbacks have generally led economists to prefer cost-based levies, which also may generate tariff revenues, over quantitative restrictions.

Exchange Controls Exchange controls regulate the rights of residents to use foreign currencies and hold offshore or onshore foreign currency deposits. They also regulate the rights of nonresidents to hold domestic currency deposits onshore. In addition, they may be defined to include taxes on currency transactions and multiple exchange rate practices that are aimed at influencing the volume and composition of foreign currency transactions. Exchange controls are not necessarily aimed at restricting capital flows; they are occasionally intended only to restrict the current account (trade in goods and services). Yet, strictly speaking, currencies are simply another type of financial asset, and therefore these controls amount to restrictions on trade in assets. Furthermore, whatever their original intention, exchange controls generally have a significant impact on the capital account.

Data on Capital Controls The primary source for internationally comparable data on capital controls is the IMF *Annual Report on Exchange Arrangements and Exchange Restrictions* (AREAER). The publication contains a detailed description of the legal framework that governs the capital account and is published annually, providing data relating to a large number of countries. Prior to 1996 the IMF reported only whether the country had imposed restrictions; after 1996 the publication includes a much more detailed description of the legal regime governing the capital account.

A number of researchers have devised various numerical indicators of the degree of capital account openness/controls, using the IMF AREAER dataset in addition to primary country sources (e.g., Miniane 2004; Edwards 2006). Several researchers alternatively focus on stock markets and detail various aspects of their actual or de jure state, such as when stock markets are open to trading by foreigners, or when domestic companies are allowed to cross-list abroad (e.g., Henry 2003; Edison and Warnock 2003; Bekaert et al. 2005).

This research agenda is still in its infancy, with little agreement among researchers on the appropriate measures and wide agreement on their respective drawbacks. As data quality improves economists should be better able to analyze and distinguish among the effects of various types of controls.

See also asymmetric information; balance of payments; Bretton Woods system; capital mobility; convertibility; currency crisis; exchange rate regimes; financial crisis; financial liberalization; hot money and sudden stops; impossible trinity; International Monetary Fund (IMF); Tobin tax; Washington consensus

FURTHER READING

Ariyoshi, A., K. Habermeier, B. Laurens, I. Otker Robe, J. Canales Kriljenko, and A. Kirilenko. 2000. "Country Experiences with the Use and Liberalization of Capital Controls." Occasional Paper No. 190. Washington, DC: International Monetary Fund. Details cross country experiences with various types of capital controls.

Calvo, G., L. Leiderman, and C. Reinhart. 1996. "Inflows of Capital to Developing Countries in the 1990s." *Journal of Economic Perspectives* 10: 125 39. Discusses the macroeconomic implications of large scale capital inflows into a developing country.

Bekaert, Geert, Campbell R. Harvey, and Christian Lund blad. 2005. "Does Financial Liberalization Spur Growth?" *Journal of Financial Economics* 77 (1): 3 55.

Bird, G., and Ramkishen S. Rajan. 2000. "Restraining International Capital Flows: What Does It Mean?" *Global Economic Quarterly* 1 (2): 57 80. Describes the various types of capital and exchange controls. Surveys the main arguments for and against the use of capital controls.

Edison, Hali J., Michael W. Klein, Luca Antonio Ricci, and Torsten Sløk. 2004. "Capital Account Liberalization and Economic Performance: Survey and Synthesis." *IMF Staff Papers* 51 (2): 220 56. Surveys the literature on the economic effects of capital account openness and stock market liberalization. Also discusses the various empirical measures used to gauge the presence of con trols on capital account transactions.

Edison, Hali J., and Francis E. Warnock. 2003. "A Simple Measure of the Intensity of Capital Controls." *Journal of Empirical Finance* 10 (1 2): 81 103.

Edwards, Sebastian. 2006. "Capital Controls, Capital Flow Contractions, and Macroeconomic Vulnerability." *Journal of International Money and Finance* 26 (5): 814 40. Along with Edison and Warnock (2003), introduces new measures of the intensity of capital account open ness and/or restrictions.

Eichengreen, B., M. Mussa, G. Dell'Ariccia, E. Detragiache, and G. Milesi Ferretti. 1999. "Liberalizing Capital Movements: Some Analytical Issues." *Economics Issues No. 17* (February). Washington, DC: International Monetary Fund. Discusses the analytical issues sur rounding the imposition of capital controls.

Henry, Peter Blair. 2003. "Capital Account Liberalization, the Cost of Capital, and Economic Growth." *American Economic Review* 93 (2): 91 96. One of several research projects that estimates the economic consequences of capital controls.

Mathieson, D., and L. Rojas Suarez (1993). "Liberalization of the Capital Account: Experiences and Issues." Occa sional Paper No. 103. Washington, DC: International Monetary Fund. Discusses cross country experiences with capital account liberalization.

Miniane, Jacques. 2004. "A New Set of Measures on Capital Account Restrictions." *IMF Staff Papers* 51 (2): 276 308.

Stiglitz, Joseph E. 2000. "Capital Market Liberalization, Economic Growth, and Instability." *World Development* 28 (6): 1075 86. Offers a critical analysis of the eco nomic impact of capital account liberalization.

Wright, Mark L. J. 2006. "Private Capital Flows, Capital Controls, and Default Risk." *Journal of International Economics* 69: 120 49. Offers a theoretical examination of the conditions that make capital controls beneficial.

ILAN NOY AND RAMKISHEN S. RAJAN

■ capital flight

Today's world economy is characterized by large movements of capital across countries and regions. Cross-border financial movements include legal transactions that are duly recorded in national ac counts as well as the illicit smuggling of capital re ferred to as *capital flight*. Capital flight is the residual difference between capital inflows and recorded foreign-exchange outflows. Capital inflows consist of net external borrowing plus net foreign direct in vestment. Recorded foreign-exchange outflows comprise the current account deficit and net addi tions to reserves and related items. The difference between the two constitutes the measure of capital flight (Erbe 1985).

Measuring Capital Flight Capital flight (KF) for a country in a given year is calculated as:

$$KF = \Delta DEBT + FDI - (CA + \Delta RES)$$

where $\Delta DEBT$ is the change in total external debt outstanding, FDI is net direct foreign investment, CA is the current account deficit, and ΔRES is net additions to the stock of foreign reserves.

A number of refinements are made to the capital flight formula to obtain a measure of capital flight to take into account the various channels and factors that affect capital flight. First, as countries' debts are denominated in various currencies, the reported U.S. dollar value needs to be adjusted to take into account fluctuations in the exchange rates of these currencies against the dollar (Boyce and Ndikumana 2001). Second, trade misinvoicing constitutes an important channel of capital smuggling (Gulati 1987; Lessard and Williamson 1987). Export underinvoicing and import overinvoicing inflate the current account deficit recorded in the balance of payments, while import underinvoicing leads to understatement of the true deficit. Thus the capital flight estimate ob tained using balance of payments trade data is likely either to overstate or understate the actual volume of capital flight. The refined measure of capital flight is the following:

$$KF = \Delta ADJDEBT + FDI \\ - (CA + \Delta RES) + MISINV$$

where *ADJDEBT* is debt flows adjusted for exchange rate fluctuations and *MISINV* is net trade misinvoicing. Additionally, a measure of the opportunity cost of capital flight or the losses incurred by the country through capital flight is obtained by imputing interest earnings on capital that left the country in early years.

Significance of Capital Flight from Developing Countries The problem of capital flight from developing countries deserves serious attention for several reasons. First, capital flight reduces domestic investment directly by reducing the volume of savings channeled through the domestic financial system, hence retarding economic growth. Second, capital flight affects the government's budget balance indirectly by reducing the tax base through reduced domestic economic activity. Moreover, capital flight forces the government to increase its borrowing from abroad, which further increases the debt burden and worsens the fiscal balance. Third, capital flight is likely to have pronounced regressive effects on the distribution of wealth. The individuals who engage in capital flight generally are members of a country's economic and political elites. They take advantage of their privileged positions to acquire and channel funds abroad. The negative effects of the resulting shortages of revenue and foreign exchange, however, fall disproportionately on the shoulders of the less wealthy members of society. The regressive impact of capital flight is compounded when financial imbalances result in devaluation of the national currency because those wealthy individuals who hold the external assets are insulated from its negative effects, while the poor enjoy no such cushion. Fourth, capital flight exacerbates the resource gaps faced by developing countries and forces them to incur more debt, which worsens their international position and undermines overall economic performance.

What Causes Capital Flight? The empirical literature has identified a number of factors that are associated with high levels of capital flight (Ndikumana and Boyce 2003; Murinde, Hermes, and Lensink 1996). The single most consistent finding in empirical studies on capital flight is that the annual flows of external borrowing are strongly associated with capital flight. The causal relationship between capital flight and external debt can run both ways: that is, foreign borrowing can cause capital flight, and capital flight can lead to more foreign borrowing. Foreign borrowing causes capital flight by contributing to an increased likelihood of a debt crisis, worsening macroeconomic conditions, and causing a deterioration of the investment climate. This is referred to as *debt-driven capital flight*. Foreign borrowing may provide the resources as well as a motive for channeling private capital abroad, a phenomenon called *debt-fueled capital flight* (Boyce 1992). Capital flight also induces foreign borrowing by draining national foreign-exchange resources and forcing the government to borrow abroad.

Good economic performance, measured in terms of higher economic growth, and stable institutions are associated with lower capital flight. Strong economic growth, for example, is a signal of higher expected returns on domestic investment, which induces further domestic investment and thus reduces capital flight. High and sustained economic growth also gives confidence to domestic investors about the institutional and governance environment of the country, which encourages domestic investment while reducing incentives for capital flight. The political and institutional environment is also an important factor in capital flight. In particular, political instability and high levels of corruption encourage capital flight as savers seek to shield their wealth by investing in foreign assets.

Policy Responses to Capital Flight Private assets held abroad by residents of developing countries include both illicit and legally acquired assets. Different strategies are required to repatriate the two types of assets. Presumably, savers choose to hold legally acquired assets abroad to maximize risk-adjusted returns. These assets will be repatriated as domestic returns rise relative to foreign returns. In order to prevent further transfers of resources abroad and to entice repatriation of legally acquired assets, therefore, governments in developing countries must implement strategies to improve the domestic investment climate.

Illegally acquired assets are held abroad not so much to maximize the returns on assets as to evade the law. Owners of these assets will be enticed by higher domestic returns only if they have guarantees of immunity against prosecution for fraud and penalties for unpaid taxes. Such guarantees, however, would have perverse incentive effects by rewarding malfeasance. Alternatively, these assets could be impounded and repatriated by force. However, the identity of the owners of capital that has been illegally smuggled out of a country can be difficult to obtain from the major financial centers, which have strong customer privacy legislation. In order to recover the proceeds of capital flight, therefore, developing countries will need the cooperation of the international financial centers and their governments.

See also balance of payments; capital controls; capital flows to developing countries; foreign direct investment (FDI); international reserves; money laundering; offshore financial centers

FURTHER READING

Boyce, J. K. 1992. "The Revolving Door? External Debt and Capital Flight: A Philippine Case Study." *World Development* 20 (3): 335 49. Explains in detail the two way causation between capital flight and external borrowing.

Boyce, J. K., and L. Ndikumana. 2001. "Is Africa a Net Creditor? New Estimates of Capital Flight from Severely Indebted Sub Saharan African Countries, 1970 1996." *Journal of Development Studies* 38 (2): 27 56. Provides a detailed description of the methodology for measuring capital flight and covers a representative sample of African countries.

Erbe, S. 1985. "The Flight of Capital from Developing Countries." *Intereconomics* (November/December): 268 75. Provides a concise definition of capital flight, which was extended by subsequent studies.

Gulati, S. K. 1987. "A Note on Trade Misinvoicing." In *Capital Flight and Third World Debt*, edited by D. R. Lessard and J. Williamson. Washington, DC: Institute for International Economics, 68 78. Describes the mechanisms of trade misinvoicing, which is an important channel for capital flight.

Lessard, D. R., and J. Williamson, eds. 1987. *Capital Flight and Third World Debt.* Washington, DC: Institute for International Economics. Discusses capital flight and its relationship with external debt.

Murinde, V., N. Hermes, and R. Lensink. 1996. "Comparative Aspects of the Magnitude and Determinants of Capital Flight in Six Sub Saharan African Countries." *Saving and Development* 20 (1): 61 78. Attempts an empirical analysis of the causes of capital flight.

Ndikumana, L., and J. K. Boyce. 2003. "Public Debts and Private Assets: Explaining Capital Flight from Sub Saharan African Countries." *World Development* 31 (1): 107 30. Provides a comprehensive survey of the literature on the determinants of capital flight and investigates the causes of capital flight by exploring the role of conventional portfolio theory factors as well as institutional and policy related factors.

LÉONCE NDIKUMANA

■ capital flows to developing countries

Since the 1970s, business cycles in many developing countries have been characterized by fluctuations in international capital flows. This has been particularly true for economies integrated into world financial markets. Fluctuations in capital flows lead to variations in the availability of financing (absence or presence of credit rationing) and in length of loans (maturities). This involves short-term volatility, such as the very intense upward movement of spreads (margin over basic interest rate that reflects perceived risk) and the interruption (rationing) of capital flows observed during the Mexican (1994 95), Asian (1997 98), and Russian (1998) crises. They also involve *medium-term* cycles, as the experience of the past few decades (since the last quarter of the 20th century) indicates.

The developing world experienced two full medium-term cycles between 1970 and 2000: a boom of external financing in the 1970s, followed by a major debt crisis, mainly in Latin America, in the 1980s; and a new boom in the 1990s, followed by a sharp reduction in net flows after the Asian and Russian crises of 1997 98. By 2007, international

capital flows to developing countries had recovered, but new sources of potential procyclicality had emerged, particularly related to the explosive growth of derivatives worldwide. Derivatives, at their simplest, allow two parties to agree on a future price for a given asset, for example, a purchase of foreign exchange. Derivative contracts have become increasingly important in developing economies as instruments for firms and others to hedge risk and for international hedge funds and investment banks to speculate. Large parts of these derivative markets are not regulated, nor have existing regulations fully incorporated the risks that derivatives pose in situations of stress, when they can add to systemic risk.

During booms, developing countries that markets view as success stories are almost inevitably drawn into a capital flows boom, inducing private-sector deficits and risky balance sheets (Ffrench-Davis 2001). For example, when Mexico joined both the North American Free Trade Agreement (NAFTA) and the Organisation for Economic Co-operation and Development (OECD) in the 1990s and was seen as a "successful reformer," it initially attracted huge inflows of capital. Even countries with weak macroeconomic fundamentals, such as low current account deficits, may be drawn into the boom, however, and all countries, again with some independence from their fundamentals, will suffer sudden stops of capital flows. There is also widespread evidence that ample private capital flows encourage expansionary macroeconomic policies during booms, such as excessive growth of government spending. Ample private capital flows encourage expansionary responses by individuals (who consume more) and companies (which invest more), as well as by macroeconomic authorities. When capital inflows fall sharply or turn into outflows, consumers, companies, and governments in developing countries have to reduce their spending. Thus unstable external financing distorts incentives that both private agents and authorities face *throughout* the business cycle, inducing a procyclical behavior of economic agents *and* macroeconomic policies (Kaminsky et al. 2004) .

Although procyclicality is inherent in capital markets, domestic financial and capital account liberalization in the developing world, as well as technological developments, such as very rapid communications, have accentuated its effects. A lag in developing adequate prudential regulation and supervision frameworks increases the risks associated with financial liberalization.

The costs of such financial volatility in the developing world in terms of economic growth are high. There is now significant evidence that capital flows have not encouraged growth and rather have increased growth volatility in emerging economies (Prasad et al. 2003). Whatever the efficiency gains from financial market integration, they seem compensated by the negative effects of growth volatility. During and after financial crises, major falls in output, employment, and investment have occurred (Eichengreen 2004). Volatility in financial markets is partly transmitted to developing countries through public-sector accounts, especially through effects of the availability of financing on government spending, and of interest rates on public sector debt service payments. In commodity-dependent developing countries, links between availability of financing and commodity prices reinforce the effects on public sector accounts. The most important effects of capital account fluctuations are on private spending and balance sheets. Capital account cycles, their domestic financial multipliers, and their reflection in asset prices (such as stock markets and property prices) became an important determinant of growth volatility.

Different types of capital flows show different volatility patterns. The higher volatility of short-term capital indicates that reliance on such financing is highly risky (Rodrik and Velasco 2000), whereas the smaller volatility of foreign direct investment (FDI) vis-à-vis all forms of financial flows is considered safer. FDI also can bring valuable benefits related to technology transfers and access to management expertise and to foreign markets. Use of risk management techniques by multinationals, via derivatives, may make FDI in critical moments as volatile as traditional financial flows, however (Griffith-Jones and Dodd 2006).

Countercyclical Prudential Regulation and Supervision Managing countercyclical policies for

developing countries in the current globalized financial world is no easy task. For this, it is essential that international cooperation in the macroeconomic policy area be designed to overcome incentives and constraints. This means that the first role of international financial institutions, from the point of view of developing countries, is to mitigate the procyclical effects of financial markets and open policy space for countercyclical macroeconomic policies, that is, policies that can attenuate the economic cycle, for example, by expanding government spending when the economy is slowing down so as to accelerate recovery and by contracting government spending in boom times, to avoid overheating of the economy. This can be achieved partly by smoothing out boom-bust cycles at the source through regulation.

One of the major problems seems to be the focus of prudential regulation on microeconomic risks and the tendency to underestimate risks that have a clear *macroeconomic* origin (see BIS 2001, chap. 7). For example, in times of rapid economic growth, a portfolio of bank loans may seem very safe; when the economy slows down or goes into recession, however, that same portfolio of bank loans may become highly problematic. This dimension of changing risk through time is not usually sufficiently perceived by individual banks or even by bank regulators. The basic problem in this regard is the inability of individual financial intermediaries to internalize collective risks assumed during boom periods.

Moreover, traditional regulatory tools, including both Basel I and Basel II international standards for bank regulation on capital adequacy, have a procyclical bias. The basic problem is a system in which loan-loss provisions are tied to loan default or to short-term expectations of future loan losses. Precautionary signals may be ineffective in hampering excessive risk-taking during booms, when expectations of loan losses are low, effectively underestimating risks and the counterpart provisions for loan losses. The sharp increase in loan delinquencies during crises reduces financial institutions' capital and, hence, their lending capacity, potentially triggering a credit squeeze; this would reinforce the downswing

in economic activity and asset prices and, thus, the quality of the portfolios of financial intermediaries.

Given the central role all of these processes play in the business cycles of developing countries, and the important influence of banking regulation on credit availability in the modern economy, the crucial issue is to introduce a countercyclical element into prudential regulation and supervision. The major innovation is the Spanish system of forward-looking provisions, introduced in 2000 and later adopted by Portugal and Uruguay. According to this system, provisions are made when loans are disbursed based on the expected ("latent") losses; such latent risks are estimated on the basis of a full business cycle.

Under this system, provisions build up during economic expansions and are drawn on during downturns. Moreover, many regulatory practices aimed at correcting risky practices shift underlying risks to nonfinancial agents (e.g., companies). This is why capital account regulations aimed at avoiding inadequate maturity structure of borrowing in external markets by all domestic agents, and at avoiding currency mismatches in the portfolios of those agents operating in nontradable sectors, may be the best available option (Ocampo 2003). Also, as long as there is no international central bank that could provide unconditional official liquidity in times of crisis, international rules should continue to provide room for the use of capital account regulation by developing countries.

More broadly, the Basel II Accord to regulate banks internationally has a number of problems that require attention: it is complex where it should be simple; it is implicitly procyclical when it should be explicitly countercyclical; and although it is supposed to more accurately align regulatory capital with the risks that banks face, in the case of lending to developing countries it ignores the proven benefits of diversification. In particular, by failing to take account of the benefits of international diversification of portfolios, capital requirements for loans to developing countries will be significantly higher than is justified on the basis of the actual risks attached to such lending. There are thus fears that Basel II creates the risk of a sharp reduction in bank lending to de-

veloping countries, particularly during crises (thus enhancing the stop-go pattern of such lending).

One clear way in which Basel II could be improved to reduce these problems would be to introduce the benefits of diversification. One of the major benefits of investing in developing and emerging economies is their relatively low correlation with mature markets. This has been tested empirically using a wide variety of financial, market, and macro variables. Different simulations that compared estimated losses of portfolios that were diversified across both developed and developing countries with the losses of portfolios in developed countries only indicate that the former were from 19 to 23 percent lower (Griffith-Jones, Segoviano, and Spratt 2002). If risks are measured precisely, this should be reflected in lower capital requirements.

An additional positive effect of taking account of the benefits of diversification is that this makes capital requirements far less procyclical than otherwise. Indeed, if the benefits of diversification are incorporated, simulations show that the variance over time of capital requirements will be significantly smaller than if they are not.

Adequate Official Liquidity for Crises At the country level, central banks have acted for many decades as lenders of last resort, providing liquidity automatically to prevent financial crises and avoid their deepening when they occur. Equivalent international mechanisms are still at an embryonic stage, with International Monetary Fund (IMF) arrangements, as of 2007, providing credit only with policy conditions attached and not automatically (Ocampo 2003). Despite some moderation in this area in the early 2000s, the general trend in IMF financing was toward increased conditionality, even in the face of external shocks, including those that involve financial contagion. Enhanced provision of emergency financing at the international level in response to external shocks is essential to lowering unnecessary reduction of economic growth or recession within a country and to avoiding the spread of crises to other countries.

Between the 1980s and the early 2000s, capital account liberalization and large capital account volatility greatly increased the need for official liquidity

to deal with large reversals in capital flows. There is increasing consensus that many of the crises in emerging markets in the late 1990s and early 2000s have been triggered by self-fulfilling liquidity runs (Hausmann and Velasco 2004). Indeed, capital outflows could be provoked by many factors not related to countries' policies. The enhanced provision of emergency financing in the face of capital account crises is thus important not only to manage crises when they occur but to prevent such crises and to avert contagion (Cordella and Yeyati 2005; Griffith-Jones and Ocampo 2003).

To address this obvious need, the IMF has made efforts to improve its lending policy during capital account crises. In 1997, the Supplemental Reserve Facility was established.

The evidence that even countries with good macroeconomic fundamentals might be subject to sudden stops of external financing also gave broad support to the idea that a precautionary financial arrangement, closer to the lender-of-last-resort functions of central banks, had to be added to existing IMF facilities. In 1999 the IMF introduced the Contingent Credit Line (CCL). The facility was never used and was discontinued in November 2003. Contrary to what was desired, the potential use of the CCL was seen as an announcement of vulnerability that could harm confidence.

After the expiration of the CCL, the IMF explored other ways to achieve its basic objectives. A move in that direction was proposed by the managing director of the IMF and approved by the International Monetary and Financial Committee in April 2006. Interestingly, as the IMF recognized, by offering instant liquidity, a well-designed facility of this sort "would place a ceiling on rollover costs thus avoiding debt crises triggered by unsustainable refinancing rates, much in the same way as central banks operate in their role of lenders of last resort" (IMF 2005). Approval of such a facility within the IMF, however, seems difficult to achieve.

Implications Volatility and contagion in international financial markets increased the incidence of financial crises and growth volatility in the developing world, and reduced policy space to adopt

countercyclical macroeconomic policies. Therefore, a major task of a development-friendly international financial architecture is to mitigate procyclical effects of capital flows and open a debate about countercyclical macroeconomic policies in the developing world. To achieve these objectives, a series of useful policy instruments can be developed, including explicit introduction of countercyclical criteria in the design of prudential regulatory frameworks; designing market mechanisms that better distribute the risk faced by developing countries throughout the business cycle; and better provision of countercyclical official liquidity to deal with external shocks. Such measures would make capital flows better support development.

See also asymmetric information; balance sheet approach/effects; banking crisis; capital controls; capital flight; capital mobility; contagion; convertibility; currency crisis; financial crisis; financial liberalization; global imbalances; hot money and sudden stops; international financial architecture; International Monetary Fund (IMF); International Monetary Fund conditionality; international reserves; Latin American debt crisis; lender of last resort; twin deficits

FURTHER READING

Bank for International Settlements. 2001. *71st Annual Report* (June). Basel: BIS. Annual report by the BIS on ongoing macroeconomic and financial issues in the global economy.

Cordella, Tito, and Eduardo Levy Yeyati. 2005. "A (New) Country Insurance Facility." IMF Working Paper No. 05/23 (January). Washington, DC: International Monetary Fund. Proposes the creation of a country insurance facility to deal with liquidity runs and financial crises and how such a facility might be designed.

Eichengreen, Barry. 2004. "Global Imbalances and the Lessons of Bretton Woods." NBER Working Paper No. 10497 (May). Washington, DC: National Bureau of Economic Research. Excellent overview of the various analytical debates about the ongoing global macroeconomic imbalance.

Ffrench Davis, Ricardo. 2001. *Financial Crises in "Successful" Emerging Economies.* Washington DC: Brookings Institution Press and Economic Commission for Latin America and the Caribbean (ECLAC). Focuses on developing countries in Asia and Latin America that have been impacted by a "new variety" of crises and discusses the main factors that might increase a country's vulnerability to such crises.

Griffith Jones, Stephany, and Randall Dodd. 2006. "Report on Derivatives Markets: Stabilizing or Speculative Impact on Chile and a Comparison with Brazil." Santiago, Chile: Economic Commission for Latin America and the Caribbean (ECLAC). Available at http://www.cepal.org.ar/publicaciones/xml/2/28572/2007 282 W134 GriffithJones.pdf. Discusses the growth of the derivatives markets in Chile and Brazil and analyzes their role in the economy.

Griffith Jones, Stephany, and José Antonio Ocampo. 2003. *What Progress on International Financial Reform? Why so Limited?* Stockholm: Expert Group on Development Issues. Offers a stock take of the progress in international financial reforms and why such progress has been fairly limited to date.

Griffith Jones, Stephany, Miguel Angel Segoviano, and Stephen Spratt. 2002. "Basel II and Developing Countries: Diversification and Portfolio Effects." Capital Market Liberalization Program, December. Available at http://www.ids.ac.uk/ids/global/finance/pdfs/FINALBasel diversification2.pdf. Discusses the impact of Basel II on developing countries.

Hausmann, Ricardo, and Andrés Velasco (2004). "The Causes of Financial Crises: Moral Failure versus Market Failure." December. Available from http://ksghome.harvard.edu/~rhausma/new/causes of fin crises.pdf. Discusses reasons behind crises in many developing countries in the 1990s and early 2000s with emphasis on Latin America.

Kaminsky, Graciela L., Carmen M. Reinhart, and Carlos A. Vegh. 2004. "When it Rains, it Pours: Procyclical Capital Flows and Macroeconomic Policies." NBER Working Paper No. 10780. Using a sample of 104 countries, the paper documents some stylized facts about the nexus between net capital flows, fiscal policy, and monetary policy.

Ocampo, José Antonio. 2003. "Capital Account and Counter Cyclical Prudential Regulation in Developing Countries." In *From Capital Surges to Drought: Seeking Stability for Emerging Markets,* edited by Ricardo

Ffrench Davis and Stephany Griffith Jones. London: Palgrave Macmillan, 217 44. Highlights the role of capital account regulations and countercyclical pru dential regulation of domestic financial intermediaries in managing a country's macroeconomic risks associ ated with capital account induced boom bust cycles.

Prasad, Eswar S., Kenneth Rogoff, Shang Jin Wei, and M. Ayhan Kose. 2003. "Effects of Financial Globalization on Developing Countries: Some Empirical Evidence." IMF Occasional Paper 220. Washington, DC: IMF. Comprehensive discussion of the empirical evidence regarding the impact of financial globalization on de veloping countries in terms of economic growth and macroeconomic volatility. Also considers what countries need to do to exploit the benefits of financial globaliza tion.

Rodrik, Dani, and Andrés Velasco. 2000. "Short Term Capital Flows." In *Proceedings of the Annual World Bank Conference on Development Economics 1999.* Washing ton DC: World Bank, 59 90. Emphasizes the role of short term debt to reserves ratio as a robust predictor of financial crises and discusses ways in which countries can reduce their financial vulnerability.

STEPHANY GRIFFITH JONES

■ **capital mobility**

Capital mobility refers to the ease with which fi nancial flows can occur across national borders. High capital mobility implies that funds are transferred relatively seamlessly from one country to another. Low capital mobility implies that financial capital does not flow as easily into or out of a particular country, and that there may be barriers hindering the capital flow. In the world of international finance there are many types of financial flows, including foreign direct investment (FDI), portfolio flows, and flows processed through the banking sector. In order to capture the mobility of such a broad range of financial flows, economists use many different ways to measure capital mobility. No single measure will capture all the essential characteristics of capital mobility. The best results are probably obtained by using several measures together. This entry groups

the many types of measures into four main cate gories: arbitrage measures (for debt flows), quantity based measures, measures of equity market integra tion, and regulatory/institutional measures (figure 1).

Arbitrage Measures The first category refers to arbitrage conditions, which involve returns on debt instruments. These are largely embodied in the in terest parity conditions: the covered interest parity (CIP), the uncovered interest parity (UIP), and the real interest parity (RIP). The basic idea behind parity conditions is that in a perfectly integrated fi nancial market, investors will detect any gaps be tween the domestic currency return on a domestic asset and the domestic currency return on a foreign asset and will seek to close that gap. In other words, arbitrage should equalize the prices of identical assets traded in different markets that is, the law of one price holds and participants in the market will behave in such a way as to remove any differences in the exchange rate adjusted returns on assets in dif ferent markets. There are important differences be tween these measures: the CIP is the narrowest of measures; the UIP is a somewhat broader measure; and the RIP is the broadest of arbitrage measures. The arbitrage conditions seek to equate rates of re turns of comparable assets across different markets/ economies. If capital mobility is high, then any dif ferences in (exchange rate adjusted) rates of return will be alleviated through arbitrage in those markets. Hence, high capital mobility implies that the various interest differentials will be low; perfect capital mo bility implies a zero differential; while nonzero dif ferentials suggest that there are barriers to capital flows.

Arbitrage conditions are probably a more appro priate way of measuring integration for certain sec tors, such as banking, than for the whole economy. The perennial problem with using such arbitrage measures, especially in developing economies, is the question of what interest rate to use, and to what extent the available interest rates are comparable across countries.

Quantity-Based Measures A growing body of literature has explored quantity-based measures of financial integration. These measures provide an
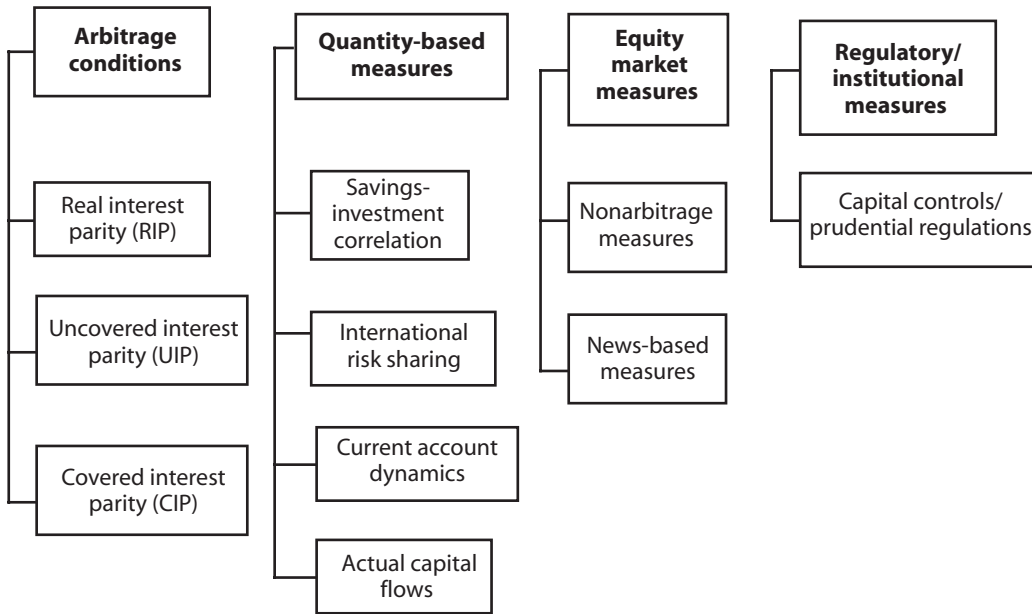
Figure 1

Categorizing measures of capital mobility: A simple framework

alternative to the traditional arbitrage conditions as a way of measuring capital mobility. Four such measures include savings-investment correlations, consumption correlations, current account dynamics, and actual capital flows.

Savings-Investment Correlations Feldstein and Horioka (1980) pioneered the use of savings-investment correlations as a measure of capital mobility. The argument regarding savings-investment correlations is that in a closed economy, by definition, savings must equal investment (i.e., the correlation between savings and investment should be very high). At the other extreme, with highly integrated capital markets and a single world interest rate, domestic investment should be largely independent of domestic savings since the former can be financed through foreign savings. There are significant empirical and theoretical shortcomings with the Feldstein-Horioka criterion and it remains a controversial measure of financial integration. The so-called Feldstein-Horioka puzzle is discussed in a separate entry.

International Risk Sharing Although savings-investment correlations are the most popular quantity measure of financial integration, a more theoretically elegant measure is examination of consumption patterns within and across economies (Obstfeld 1989). Essentially, agents access capital markets to save or dissave based on how they wish to smooth levels of consumption over time. If two agents in different countries have similar consumption patterns, this implies that they use the same capital markets and that the markets are equally accessible to both. More specifically, a high level of correlation of consumption between two economies is an indication that each country is accessing the same capital market to choose a time path for consumption that is outside the path implied by available domestic resources, thus implying access or openness to international capital flows.

The intuition behind tests of consumption correlations ("international risk sharing") is that financial openness ought to afford individuals the opportunity to smooth consumption over time as they can borrow and lend on international financial markets. Thus consumption in any one country should co-move less with income over time, and if their preferences over consumption are similar, consumption should be correlated across countries.

Conceptually, although consumption-based tests of capital mobility are attractive when attempting to discern whether a region is ready for monetary union (as the degree of business cycle synchronization may be less relevant as long as agents can share consumption risks across borders), they are based on a number of restrictive assumptions that limit their practical usefulness.

Current Account Dynamics A related strand of the literature has focused on current account dynamics and, in particular, whether the current account is stationary (i.e., its mean and variance do not change over time). Simply put, the argument here is that if savings and investment are cointegrated (i.e., a linear combination of the two is stationary), their difference, which is the current account, ought to be stationary (Ghosh 1995). The problem with this line of reasoning is that a finding of stationarity could imply either that an economy is not financially integrated (thus suggesting the existence of a long-run relationship between savings and investment) or that the open capital market is imposing a solvency constraint on the country in the sense that the financial market will penalize a country that is viewed as being profligate by persistently running current account deficits (i.e., spending more than it is producing).

Actual Capital Flows A fourth quantity-based measure of capital mobility is observation of the actual magnitude of capital flows (FDI, portfolio flows, bank flows). All other things being equal, the higher the levels of capital inflows and outflows, the greater the level of capital market integration. Although examination of cross-border capital flows is useful in examining the composition of flows, insofar as there is no yardstick by which to gauge high versus low capital mobility, it tends to be of limited use as a measure of financial integration.

Measures of Equity Market Integration Another measure of the integration of international capital markets involves examining equity market returns. This measure essentially refers to those nonarbitrage price-based measures that include stock market correlations (both direct correlations as well as the extent to which risky assets can be priced using the international capital asset pricing model), or news-based measures (i.e., the extent to which interest rates and other financial market variables are affected by common shocks versus country-specific ones).

Papers measuring nonarbitrage measures examine the bivariate properties of the data and how movements in the equity markets in one country influence the series in another country. In general, the methodological applications range from simple correlations and covariances to value-at-risk-based approaches such as Granger causality and variance decompositions for the short-run analysis and vector error correction models and cointegration tests for the long-run scenario. Essentially, these approaches examine the effect of changes in stock markets in major financial centers (the United States, Japan, the United Kingdom, etc.) on local stock markets. In addition, the use of test of asset pricing models has gained popularity. Asset pricing models allow for risk characteristics to be considered when evaluating market data for different countries. The variance of stock returns provides very useful information about the extent of capital mobility, in that a variance that decreases as the international portfolio increases would imply that correlations between markets are low enough for benefits to diversification to be realized. This is evidence against integration between these markets.

The news-based measures tend to be analyzed in a multivariate setting. The objective is to test for the existence of common trends and common sources of variation among a group of markets. A single common trend, for instance, implies a high level of integration. The existence of country-specific sources of variation would imply the opposite.

Regulatory, Institutional, and Other Measures The degree of capital mobility can also be measured by observing the extent to which a country has imposed capital controls. The types of controls that might be in place are numerous, including legislative control over deposit rates, restrictions on capital account transactions such as restrictions on term or currency, regulations relating to the entry and exit of foreign financial services, and exchange controls. An obvious limitation of these measures is the difficulty

of obtaining good proxies to measure such barriers or regulatory impediments that prevent financial integration. A fundamental assumption with all indexes of capital controls is that the removal of capital controls may, in some way, result in a more financially integrated economy. A situation could exist, however, in which a country has very few capital controls and yet is not regarded as being integrated with other economies. This could be due to legal or political factors, cultural variables, business practices, or the economy's simply not having been noticed by others as a potential place to export capital flow that is, it "escaped the radar" of the international financial community.

Going Forward It is generally believed that innovations in communications and market access, including reduction in barriers to capital flows, have increased the extent of international capital mobility worldwide. Obtaining empirical evidence of this is difficult, however. There is no single measure of capital mobility. Each measure offers only a partial indication of the extent of integration. Given the multiplicity of definitions of capital mobility, an important area for future research would be to develop a multivariate methodology to reduce the multidimensionality of the concept of capital mobility to an operational univariate measure. It is equally important to undertake more detailed studies on the legal, institutional, and other barriers that hinder the free movement of cross-border capital.

See also balance of payments; capital controls; capital flows to developing countries; convertibility; exchange rate forecasting; exchange rate volatility; Feldstein-Horioka puzzle; home country bias; interest parity conditions; purchasing power parity

FURTHER READING

Bayoumi, T. 1997. *Financial Integration and Real Activity*. Manchester, UK: Manchester University Press. Good reference for several measures of financial integration.

Feldstein, M., and C. Horioka. 1980. "Domestic Saving and International Capital Flows." *Economic Journal* 90: 314 29. The seminal reference for the literature on saving/investment correlations.

Flood, R., and A. Rose. 2003. "Financial Integration: A New Methodology and an Illustration." NBER Working Paper No. 9880. Washington, DC: National Bureau of Economic Research. A recent extension of the equity market integration literature using a general intertemporal asset pricing model.

Ghosh, A. 1995. "International Capital Mobility amongst the Major Industrialized Countries: Too Little or Too Much?" *Economic Journal* 105: 107 28. A good reference on theoretical and empirical underpinnings for consumption correlations.

Obstfeld, M. 1986. "Capital Mobility in the World Economy: Theory and Measurement." *Carnegie Rochester Conference Series on Public Policy* 24: 55 104. A good reference on theoretical and empirical underpinnings for consumption correlations.

. 1989. "How Integrated Are World Capital Markets? Some New Tests." In *Debt, Stabilization, and Development*, edited by G. Calvo. Oxford: Basil Blackwell, 134 55. A reading on consumption correlations.

Phylaktis, J., and F. Ravazzolo. 2005. "Stock Market Linkages in Emerging Markets: Implications for International Portfolio Diversification." *Journal of International Financial Markets, Institutions, and Money* 15: 91 106. A useful reference for the more recent financial integration literature involving equity markets.

TONY CAVOLI AND RAMKISHEN S. RAJAN

■ carry trade

The term *carry trade*, or *currency carry trade*, refers to trades in which funds are borrowed in a relatively low-yielding currency to invest in a higher-yielding currency. In the process, a "short" position is established in the low-yielding currency and a "long" position is created in the other currency. The degree of leverage depends on the amount of capital (or the required margins) put up against the borrowing in the relatively low-yielding currency. The borrower benefits not only from the difference in yields but, more important, from any favorable change in the exchange rate. In principle, carry trades can take place between any assets that offer different expected rates of return. This entry covers carry trades in currencies.

The considerable recent attention paid to carry trades reflects the judgment of some market analysts that these trades can significantly influence exchange rate dynamics. Surprisingly, however, there is little quantitative evidence on their importance. This, in large measure, reflects the opaqueness of the over-the-counter (OTC) markets in which the bulk of currency trading takes place, along with the difficulties of identifying the size of such trades from balance of payments data. Much of the evidence on carry trades tends to be anecdotal and based on the assessments of currency market analysts or participants.

The possibility of systematically earning higher-than-normal returns from carry trades is closely linked to whether foreign exchange markets are informationally efficient. In an efficient market, such trades would not be expected to systematically produce above-average returns (adjusted for risk), and any abnormal returns would be a matter of chance. Either because foreign exchange markets are not informationally efficient or on account of differences in view about the appropriate underlying equilibrium model of returns, the carry trade literature assumes that market players seek to systematically earn superior returns.

Market Efficiency The *informational efficiency* of financial markets refers to whether asset prices fully reflect all available information. The concepts of informational efficiency and rational expectations are closely related, and the term *rational expectations* is also used to characterize situations in which expectations (and markets) are informationally efficient.

Since the move toward generalized floating in the early 1970s there has been an extraordinary amount of empirical work on the efficiency of major foreign exchange markets. For the most part, the work has not been successful in unambiguously determining whether major foreign exchange markets are informationally efficient. Largely because efficiency tests are joint tests of both informational efficiency and an assumed underlying model of equilibrium returns, statistical findings invariably are open to different interpretations. As applied to the foreign exchange market, a common assumption in many tests has been that interest-earning assets denominated in different currencies are perfect substitutes and carry

the same expected rate of return (uncovered interest rate parity). Based on this assumption together with the assumption of covered interest rate parity, which allows the forward rate to be used as a proxy for the unobserved expected future spot rate many early tests considered whether forward exchange rates were biased predictors of future spot exchange rates (see equation [1]) or whether the difference between the realized spot rate and the lagged forward rate could be predicted on the basis of the available information set $I(t)$ (see equation [2]).

Under the null hypothesis of market efficiency, the parameter α in equation (1) should be equal to zero and the parameter β equal to unity, while the error term should be white noise. Alternatively, informational efficiency can be considered with relation to equations such as (2) where depending on the assumed content of the information set $I(t)$ the market is either weakly or strongly efficient when $\Gamma = 0$ and the error term is white noise. Here lower case s and f refer to the natural logarithms of the spot and forward rates, respectively, $I(t)$ is the information set at time t, υ and η are error terms, and Δ is the first difference operator. The term $f(t, k)$ refers to the period t forward rate for period k.

$$\Delta s(t + k, k) = \alpha + \beta(f(t, k) \\ - s(t)) + v(t + k) \qquad (1)$$
$$s(t + k) - f(t, k) = \Gamma I(t) + \eta(t + k) \qquad (2)$$

Almost without exception, tests based on equations such as (1) and (2) have rejected the informational efficiency of foreign exchange markets under the maintained assumption of interest rate parity. In particular, a relatively robust finding across many major currencies and time periods sometimes referred as to the negative forward bias puzzle is that β in equation (1) is closer to negative unity than to positive unity as implied under efficiency. In addition, it is not uncommon for available information including lagged values of forward and spot exchange rates to be found useful in helping predict the difference between forward and spot exchange rates in equations such as (2). In short, forward premiums appear to be biased predictors of future exchange rate changes and forecast errors tend to be predictable.

These findings are open to a number of different interpretations with regard to whether it is informational efficiency or uncovered interest rate parity that is rejected. Hence the underlying equilibrium model of returns is not conclusive. This issue has not been resolved, but recent tests of efficiency have increasingly moved beyond the assumption of uncovered interest rate parity to incorporate both constant and time-varying risk premiums. In addition, a number of studies have used survey data to measure exchange rate expectations in place of forward rate data.

Carry Trades and Exchange Rate Regimes The extent to which large carry trades can be built up is likely to be influenced importantly by the exchange rate regime. Under a pure floating regime, uniformity in exchange rate expectations across agents would be expected to exclude the possibility of a very large buildup in carry positions. Negative sentiment toward a currency would be expected under floating rates to be reflected relatively quickly in adjustments in interest rates and exchange rates as individuals sought to build up their short-long currency positions. As a result, risk-adjusted expected returns on assets denominated in different currencies would tend to be equalized. Any substantial buildup in carry positions under flexible exchange rates could arise when agents hold very different exchange rate expectations or there are differences in risk appetites. In such circumstances, one individual's "long" carry position on a currency would be matched by an equal and offsetting position on the part of other individuals.

When there is substantial exchange market intervention and sterilization, the emergence of large carry positions seems more plausible. Such positions would represent bets of the private sector vis-à-vis the official sector about the future time path of the exchange rate. The positions would be facilitated by the official sector intervening in the foreign exchange market and providing increased supplies of the currency in which the private sector sought to go "long." Of particular interest in this latter connection are situations in which the buildup in "long" positions in a currency is matched by the authorities' sterilization of their exchange market interventions. In these cases, the long carry position of the private sector would effectively be matched by the authorities assuming the short position in the currency.

Carry Trades and Asia Notwithstanding the large amount of attention paid to carry trades, the evidence of their importance in particular episodes is largely in the eye of the beholder. In analyzing the 1997 98 Asian financial crisis, international organizations such as the International Monetary Fund took the view (based largely on market intelligence) that carry trades had been important determinants of exchange market dynamics. Spurred by sterilized intervention that led to interest rates in many Asian emerging markets remaining well above those in the United States and Japan even as currencies were expected to appreciate, carry trades were seen as contributing importantly to the surge in short-term capital flows to the region before the crisis. Subsequently, the unwinding of these trades during the 1997 98 crisis was seen as contributing to significant downward pressure on many regional currencies. No attempt was made, however, to quantify the importance of these trades.

More recently, many market analysts expressed concern about a possible disorderly appreciation of the Japanese yen during 2006 related to a possible earlier buildup of large carry trades that had been funded in yen during the period when short-term interest rates in Japan were effectively zero. As the Bank of Japan began to raise interest rates in 2006, the possibility that these trades might be unwound led to concerns that the Japanese yen would appreciate sharply. In the event, the yen continued to remain weak even as short-term interest rates in Japan were increased modestly during the first half of 2006. Subsequently, there has been debate among private sector financial analysts as to how large the buildup in yen-funded carry trades during the zero interest-rate period had been and whether its importance had been exaggerated.

Carry trades can be viewed as one of a large number of ways in which individuals can speculate on currency movements. The absence of data on the quantitative significance of carry trades precludes a systematic evaluation of their potential importance, but many market analysts regard them as potentially

important influences on exchange markets dynamics during particular episodes.

See also Bank of Japan; capital flows to developing countries; conflicted virtue; currency crisis; exchange rate regimes; exchange rate volatility; foreign exchange intervention; hedging; interest parity conditions; international reserves; peso problem; speculation; sterilization

FURTHER READING

Anderson, Jonathan. 2006. *Japan, Free Money, and the World: Asian Economic Perspectives*. Hong Kong: UBS Investment Research. Provides a critical review of the role of the yen carry trade in 2005 and 2006 and concludes that its significance was limited.

Bank for International Settlements. 2007. "International Banking and Financial Market Developments." *BIS Quarterly Review* Part 4 (December). Basel. Discusses the role of the carry trade and its potential significance for exchange market dynamics.

Fama, E. F. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance* 25: 383 417. The definitive original overview of efficient capital markets theory and the early empirical evidence.

Hallwood, C. Paul, and Ronald MacDonald. 2000. "Spot and Forward Exchange Rates and the Efficient Market Hypothesis." In *International Money and Finance*, edited by C. Paul Hallwood and Ronald MacDonald. Malden, MA: Blackwell, chapter 12. A comprehensive survey of the literature on the efficiency of the foreign exchange market.

International Monetary Fund. 1998. *International Capital Markets: Developments, Prospects, and Key Policy Issues*. World Economic and Financial Surveys. Washington, DC: International Monetary Fund. Discusses the role of the carry trade in the lead up to the Asian financial crisis and in the subsequent crisis dynamics.

CHARLES ADAMS

■ **Central American Common Market (CACM)**
The Central American Common Market (CACM) was established by the General Treaty of Central American Economic Integration signed on December 13, 1960, by Guatemala, Honduras, El Salvador,

and Nicaragua. Costa Rica joined on July 23, 1962. The treaty was the principal economic component of a series of measures adopted during the 1950s and early 1960s that were designed to facilitate the political and economic integration of the five nations of the Central American isthmus. The General Treaty also created two regional institutions: the Secretariat for Regional Economic Integration and the Central American Bank for Economic Integration. The political component of this integration process was to be carried out through the Organization of Central American States (ODECA), which was formed on October 14, 1951.

Integration Efforts in the 1950s and 1960s The notion of a unified isthmus has a long history in Central America. It dates back at least to 1821 and the formation of the Federation of Central American States shortly after the region's independence from Spain. The federation lasted only 15 years, but the ideal of unification continued to prevail throughout the next century and finally led to concrete unification measures in the 1950s. During this period, regional integration was driven in part by the example of the European Community and by the views of economists at the UN Economic Commission for Latin America (ECLA). According to their import-substitution industrialization (ISI) strategy, poor nations should eschew international trade and turn inward by closing their internal market to foreign products. The resulting boost to their domestic industry would lead to higher living standards. Since the ISI strategy does depend on the size of internal markets, however, ECLA economists believed that the strategy would be more effective in the region if the five nations formed a free trade area that expanded the market size for local firms but still protected them from international competition. Prompted by these ideas, the five nations signed the Multilateral Treaty on Free Trade and Economic Integration in Central America on June 10, 1958. That same day, they also signed the Regime for Central American Integration Industries (RII), which permitted member countries to identify firms that due to economies of scale would be granted monopoly right to supply the entire region. To

ensure balanced development, each nation would be allocated an equal number of these industries.

The ECLA approach led to opposition from the United States and other groups concerned with its protectionist bent. Responding to this pressure, the five nations reached a new agreement the General Treaty that altered the direction of the proposed common market from one focused on protection to one intended to be more encouraging of trade. For instance, whereas the Multilateral Treaty required that products eligible for free trade intraregionally be listed within the treaty, the General Treaty freed all products unless specifically exempted and identified by the treaty.

By most criteria, the CACM was considered a dramatic success, especially during its first decade of operation. According to Bulmer-Thomas (1998, 314 16), the elimination of duties on intraregional trade and the creation of a common external tariff (CET) was done rapidly and efficiently: 74 percent of goods listed on the tariff schedule traded freely immediately on the General Treaty's entry into force and, by the end of 1966, 94 percent of all listed products were assessed no duty when trading intraregionally. In addition, by 1967, 90 percent of traded goods were covered by the CET. As a consequence, intraregional trade expanded by a factor of nine, from $31.3 million in 1960 to $285.2 million in 1970, and the share of intraregional trade as a percentage of total exports rose from 7.0 percent in 1960 to 26 percent in 1970. For more details see Bulmer-Thomas (1998, table 1).

Despite these successes, deficiencies in the integration model soon became evident. The General Treaty was biased toward freeing up intraregional trade in consumer goods manufactured regionally and left in place barriers to intraregional agricultural trade. This contributed to sizable intraregional trade imbalances among the less-developed members such as Honduras and Nicaragua, which had small or nonexistent manufacturing sectors. It also led to sizable trade diversion as third-country manufactured consumer goods, which faced high protectionist CETs, were replaced by more expensive, lower-quality regional goods. Yet the mechanisms

established to address imbalances were either incapable of reaching a consensus or nonexistent. For instance, the RII was suspended when Costa Rica permitted the establishment of a tire factory to compete with an integration industry already operating in Guatemala. Finally, since the structure of the regional agreement still maintained elements of the ISI model, tariff revenues fell sharply as a percentage of total government revenues. Not only did consumer goods trade duty-free, since they were manufactured regionally, but also imports of intermediate and capital goods from outside the region entered duty-free to avoid harming these regional manufactures. To address this problem, the five nations agreed to a 30 percent increase in the CET under the San Jose Protocol of 1968.

Finally, regional integration collapsed when a dispute over the attempted expulsion by Honduras of Salvadoran immigrants during the summer of 1969 led to a four-day war between the two neighbors and the eventual withdrawal of Honduras from the CACM in December 1970. Relations between the two nations would remain suspended for a decade.

A Revival of Integration Efforts in the 1990s As a by-product of efforts to restore peace to Central America, the Central American presidents' meeting in Antigua, Guatemala, in 1990 called for reviving the integration process. In response, the five nations, including Panama as a new member, signed the Tegucigalpa Protocol to the ODECA Charter on September 13, 1991. The protocol amended the existing integration framework and established a new institution, the System of Central American Integration (SICA), which was to oversee the integration process. As an umbrella institution, SICA oversaw 4 thematic areas (economic, political, social, and environmental) and incorporated 27 other regional institutions, more than 200 treaties or protocols existing among the member nations, and 3 regional bodies: the Central American Court of Justice, the Central American Parliaments, and the Secretary General of SICA.

It is worth underscoring that the protocol called for the creation of an economic union and the positioning of the region within the global economy.

This was one of a number of important indicators that views within the region regarding the appropriate development model had shifted dramatically. The integration efforts were now to be guided by an export-led development model whose orientation was outward, as opposed to the inward-oriented, import-substitution framework that dominated the views of the early architects of the CACM. Regionalism was now to be open, not closed. Moreover, several of the member countries had adopted a neoliberal trade strategy that focused on lowering tariffs, eliminating quantitative trade barriers, and employing more market-friendly measures.

The next important step in the revival of economic integration was taken on October 29, 1993, when the six member nations of SICA signed the Guatemala Protocol to the General Treaty. This protocol focused on economic integration and committed the six nations to the formation of a customs union.

Since these steps were taken, the momentum toward integration has ebbed and flowed, buffeted by a host of factors internal (such as the election of presidential candidates opposed to further integration) and external to the region. For instance, Hurricane Mitch in 1998 dampened enthusiasm for integration, the prices of the region's commodity exports deteriorated sharply, and SICA was confronted with a severe funding crisis.

Nonetheless some progress has occurred and the enthusiasm for integration appeared to be on the rise by about 2005. By early 2007, all but a handful of regional goods traded freely within the region and, according to data provided by SICA, 94 percent (5,846 out of 6,198) of the products in the tariff schedule had been harmonized into a CET. Additionally, the CET was far less protectionist than during the earlier integration period and was applied as follows: 0 percent to capital goods and raw materials produced outside the region, 5 percent to raw materials produced regionally, 10 percent to intermediate goods produced regionally, and 15 percent to final goods. Moreover, the recent implementation of the Central American–Dominican Republic Free Trade Agreement (CAFTA-DR), which establishes a free trade area (FTA) that includes the United States, the five Central American countries, and the Dominican Republic, has great potential for deepening the region's integration.

Following a comprehensive comparison of the FTA with existing integration instruments, Gonzalez (2005) concludes that among the most important features of the new FTA is that the Central American countries opted to make its requirements apply in a plurilateral fashion, as opposed to only bilaterally between the United States and each country. This contrasts with the FTA agreements signed earlier by the region with Chile and with Mexico, which were hub-and-spoke in nature. A second important feature identified by Gonzalez is that the FTA agreement sets a "floor" on disciplines in a host of existing areas (e.g., trade in goods) and new areas (e.g., services, investment, intellectual property, dispute settlement, etc.). As a result, the new FTA essentially deepens and updates the regional integration instruments.

See also Central American Dominican Republic Free Trade Area (CAFTA-DR); common market; free trade area; Free Trade Area of the Americas (FTAA); import substitution industrialization; regionalism; tariff rate quotas

FURTHER READING

Bulmer Thomas, Victor. 1998. "The Central American Common Market: From Closed to Open Regionalism." *World Development* 26 (2): 313 22. Contains a useful bibliography and updated analysis of the new integration efforts by the leading scholar on economic integration in the region.

Cline, William R., and Enrique Delgado, eds. 1978. *Eco nomic Integration in Central America*. Washington, DC: Brookings Institution. Excellent, comprehensive analy sis of the initial period of integration.

Gonzalez, Anabel. 2005. *The Application of the Dominican Republic Central America United States Free Trade Agreement*. OAS Trade, Growth, and Competitiveness Studies. Washington, DC: Organization of American States. Principally a legal analysis, but also provides ex cellent insights on the economic implications of CAFTA DR for regional integration.

Solis, Luis G., and Patricia Solano. 2001. *Central America: The Difficult Road towards Integration and the Role of*

Canada. Ontario, Canada: Canadian Foundation for the Americas. Provides a useful overview of SICA's early history and outlines measures Canada may adopt to support integration efforts in the region.

JOSE A. MENDEZ

■ Central American–Dominican Republic Free Trade Area (CAFTA-DR)

On May 28, 2004, after a year and a half of intense negotiations, the United States and five Central American countries (Costa Rica, El Salvador, Guatemala, Honduras, and Nicaragua) signed the Central American Free Trade Agreement (CAFTA). The Dominican Republic joined the agreement on August 5, 2004, and to reflect this, the acronym was changed to CAFTA-DR. Like other recent free trade agreements (FTAs) signed by the United States, CAFTA-DR is a comprehensive agreement that goes far beyond traditional FTAs. Once fully implemented the agreement will bring about the elimination of barriers to virtually all trade and investment, and each member country will have implemented legal and regulatory reforms designed to protect intellectual property, raise and enforce labor and environmental standards, improve customs administration, and open up government procurement.

By December 2007, the treaty had been implemented between the United States and four countries and implementation with the Dominican Republic was imminent. Costa Rica had yet to submit the treaty to its congress for a vote, but ratification was certain following approval on October 7, 2007 by 51.6 percent of Costa Rican voters of a referendum supporting the treaty.

Challenges and Opportunities For the Central American (CA) countries, the FTA offered an opportunity to enhance their protrade, market-oriented development strategy by locking in the strategy with a trade agreement with the United States, their principal trading partner as well as the world's largest market. In 2004, the United States accounted for 56 percent of the region's exports and 44 percent of the region's imports. The agreement would also secure access on a reciprocal and more permanent basis. Eighty percent of the region's exports to the United States already entered duty free, but under the Caribbean Basin Initiative and General System of Preferences, tariff programs are one way, highly limited, and subject to periodic review and approval by the U.S. Congress.

U.S. policymakers saw an opportunity to improve national security because boosting regional prosperity would strengthen the region's ability to cooperate on security. They also saw an opportunity to pry open global markets for U.S. exporters and investors. The U.S. Trade Representative (USTR) notes that in 2004 the region represented the 2nd largest U.S. export market in Latin America behind Mexico and the 14th largest worldwide, ahead of India, Indonesia, and Russia combined.

The treaty posed two major challenges. First, it had to be perceived as balanced and not one in which the United States used its large economic size to gain the bulk of concessions. In 2004, the six countries had a combined population of about 45 million and a joint gross domestic product (GDP) of $90.7 billion, whereas the U.S. population stood at 294 million and GDP amounted to $11.7 trillion. Second, the treaty needed to balance the beneficial competitive pressures that come from freer trade with a recognition that the burden of adjustment would fall heaviest on the Central American countries since their economies were relatively more closed. Their average rate of protection was estimated to be more than three times that for the United States. Also, a high share of their economy was vulnerable. In 2004, the trade exposure index (the ratio of exports and imports to GDP) ranged from a low of 49.4 percent for Guatemala to a high of 94.3 and 95.8 percent, respectively, for the Dominican Republic and Costa Rica. Yet the Central American countries were least able to counteract the harmful adjustment effects. In 2004, per capita incomes (valued at purchasing power parity, or PPP) ranged from a high of $9,887 for Costa Rica to a low of $2,677 for Nicaragua, whereas the U.S. per capita GDP (valued at PPP) was $39,710.

Agreement Highlights The treaty commits the signatories to the elimination of import tariffs on nearly all goods traded among member countries. The only products exempted from tariff reductions are imports of sugar by the United States, white maize by El Salvador, Guatemala, Honduras, and Nicaragua, and potatoes and onions by Costa Rica.

The majority of products will receive zero-duty status immediately on the treaty's entry into force. According to estimates provided by USTR (2005), nearly 80 percent of U.S. exports of consumer and industrial goods to the region and 50 percent of its agricultural exports will receive immediate duty-free access. The tariffs on the remaining exports of industrial goods would be phased out in 10 years, whereas those facing U.S. agricultural products would be subject to tariff phase-out periods ranging from 5 to 20 years. For regional exporters, nearly 100 percent of their exports to the United States will receive immediate duty-free access, with only 19 products restricted to a 10-year phase-out period.

In addition to the lengthy transition provisions for duty reductions in order to lessen adjustment effects, the treaty permits tariff-rate quotas (TRQs) to be established for sensitive agricultural products. TRQs provide immediate zero-duty access to the quota amount of imports, but above-quota imports are assessed a higher, prohibitive rate. A special temporary safeguard rule was also established to prevent harm to domestic agriculture during the transition.

The United States made two important concessions in terms of market access. First, it agreed to double the region's sugar quota allocation over a 15-year period. Second, it relaxed the rules of origin pertaining to apparel, allowing certain apparel to enter duty-free even if the fabric was not produced in the United States. Eligible textile and apparel imports would also receive duty-free treatment retroactively to January 1, 2004.

In the services area, the treaty commitments go beyond those under the World Trade Organization's General Agreement on Trade in Services. Most industries receive broad market access. United States negotiators were also successful in opening access to sensitive sectors such as the telecommunications and insurance industries in Costa Rica. By late 2007 all that remained was passage of the necessary implementing legislation, considered a virtual certainty given the composition of the legislature. The related investment provision locks in rights for foreign investors that are already recognized in the region, such as the right to national treatment and nondiscrimination, fair compensation for expropriation, free transfer of profits, third-party arbitration, etc. A unique feature of this FTA is that it provides for a tribunal to review decisions by arbitration panels to ensure that foreign investors are not using them to circumvent domestic judiciaries (World Bank 2006).

The treaty includes detailed rules for improvement of customs procedures, enhancing transparency and access to government procurement bidding, and the monitoring and enforcement of labor and the environment standards. Unique aspects of the treaty related to the latter include: (1) establishing the right of citizens to request investigations if they believe environmental laws have been violated, (2) a commitment from the United States to strengthen the region's capacity to monitor labor and environment standards, and (3) the establishment of fines of up to $15 million for each instance of nonenforcement of labor laws.

Several studies have developed estimates of the likely economic effects on member countries of forming CAFTA-DR. Despite differences in techniques and scope, the conclusions are broadly similar. For the United States, they show that the aggregate economywide effect will be positive, but negligible. According to the largest estimate of the economic effects, by Brown et al. (2005), liberalizing trade in goods improves U.S. economic welfare by 0.04 percent of U.S. gross national product (GNP), while liberalization in services adds an additional 0.13 percent of U.S. GNP. The remaining studies, though limited to the effects of removing barriers to trade in goods, project a positive and even smaller impact on the U.S. economy. Two other studies (surveyed in Brown et al. 2005) project an improvement of GDP of 0.02 and 0.01 percent, whereas USITC (2004) obtains a positive impact on

U.S. welfare of $166 million which, when rounded, amounts to 0.00 percent of GDP.

The consequences for the regional economies are projected to be more significant. Brown et al. estimate an increase of 4.4 percent in Central American countries' GNP, whereas the other two studies calculate a rise of 2.4 and 1.5 percent in GDP.

It is important to note that the projections just cited for the Central American countries may be viewed as "minimum" estimates since they do not take into account the likely impact on foreign investment flows. Nor do they take into account the effects of locking in access to the U.S. market and providing a more stable regulatory environment. On the other hand, as underscored by the World Bank (2006), for these gains to be fully realized and spread equitably, the Central American countries will need to make complementary investments in areas such as infrastructure and create programs to assist the most vulnerable groups, that is, the poor, so they "have the means to take full advantage of the new opportunities."

See also Central American Common Market (CACM); free trade area; Free Trade Area of the Americas (FTAA); regionalism; tariff rate quotas

FURTHER READING

Brown, Drusilla K., Kozo Kiyota, and Robert M. Stern. 2005. "Computational Analysis of the U.S. FTAs with Central America, Australia and Morocco." *World Economy* 28 (10): 1441 90. Contains a useful outline of CAFTA's principal features and a valuable comparison of their own computations of CAFTA's economic impact to those developed by other authors.

Hornbeck, J. F. 2005. "The Dominican Republic Central America United States Free Trade Agreement (DR CAFTA)." CRS Report RL31870. Washington, DC: Congressional Research Service. Exhaustive, but excellent, description of CAFTA DR, including its history and the motives or issues driving the negotiators.

Salazar, Jose M., and Jaime Granados. 2004. "The U.S. Central America Free Trade Agreement: Opportunities and Challenges." In *Free Trade Agreements, U.S. Strategies and Priorities*, edited by Jeffrey J. Schott. Washington, D.C.: Institute for International Economics,

225 75. Another thorough review of CAFTA DR that identifies the key actors and their interests.

USITC. 2004. "U.S. Central America Dominican Republic Free Trade Agreement: Potential Economywide and Selected Sectoral Effects." Investigation No. 2104 13, USITC Publication 3717. Washington, DC: United States International Trade Commission.

USTR. 2005. "Free Trade With Central America and the Dominican Republic, Highlights of the CAFTA." *CAFTA Facts*. Washington, DC: Office of the United States Trade Representative.

World Bank, Latin America and the Caribbean Region. 2006. "DR CAFTA: Challenges and Opportunities for Central America." Washington, DC: World Bank. Especially valuable for its analysis of CAFTA DR's likely impact on the poor and its thoughtful assessment of the potential for trade policy changes to spur economic development.

JOSE A. MENDEZ

■ child labor

There are an estimated 191 million economically active children ages 5 to 14 in the world today (ILO 2006). This corresponds to 16 percent of this age group. The images that pervade the popular press of children chained in factories, forced into prostitution, or coerced into a country's military do not represent the conditions of most working children around the world. Most working children are at their parents' sides, helping in the family farm or business. A 2000 UNICEF (United Nations Children's Fund) project surveyed working children in 36 developing countries. The data represent more than 120 million children ages 5 to 14. Although nearly 70 percent of children in these countries spend time in some form of economic activity or domestic chores, less than 3 percent work in the formal wage labor market. Most of this wage employment, like most employment overall in the world's poorest economies, is in agriculture.

What Is Child Labor? There is no universally accepted definition of child labor. Some researchers view the phrase as referencing all nonschool, non-

leisure activities of children. Others define "child labor" as referring to activities that harm the child in some sense.

Most quoted estimates of the incidence of child labor come from the International Labor Organization's (ILO) Statistical Information and Monitoring Program on Child Labor (SIMPOC). In their most recent global estimates of child labor, they define child labor as:

- An economically active child under 12 who works 1 or more hours per week;
- An economically active child 14 and under who works at least 14 hours per week or 1 or more hours per week in activities that are classified as a worst form of child labor;
- An economically active child 17 and under who works more than 43 hours per week or 1 or more hours per week in a worst form of child labor.

SIMPOC estimates that there are 218 million child laborers in 2004 under this definition (ILO 2006). This number is larger than the 191 million economically active children ages 5 to 14 because of older children in worst forms of child labor or working intensively.

This reference to *activities classified as a worst form of child labor* refers to activities that are considered either unconditional worst forms of child labor or hazardous work. *Unconditional worst forms of child labor* are activities that ILO convention C182 on the Worst Forms of Child Labor lists as inappropriate for children under 18: forced and bonded labor, prostitution, pornography, illicit activities, soldiering, and child trafficking. Each signatory country of C182 defines a set of industries and occupations that are classified as *hazardous work* and therefore a worst form of child labor. Hazardous work is defined as economic activity that, owing to the nature or circumstance of the work, is harmful to the child. There are an estimated 126 million children in hazardous work worldwide (ILO 2006), and 8.4 million children are involved in unconditional worst forms of child labor, 68 percent of whom are in forced or bonded labor (ILO 2002). A common bonded labor arrangement is when parents are paid an advance on the child's future

labor earnings, and the child is then committed to the employer until the advance is repaid.

The Child Labor Decision Poor families balance the child's potential economic contribution against alternative uses of child time. What is the child's potential economic contribution? Direct wage income paid to working children may be important in some contexts, but wage work is rare. In most contexts, working children's primary economic contribution comes through the help they offer their families. Most often, this help is providing domestic services that free up adult time for income-generating pursuits. When there is a family business or farm, the child and other family members often help, and working in the family business or farm is the most prevalent economic activity of children. The value of the child's economic contribution to family farms and businesses can be large. One recent study from Nepal estimates that children are responsible for nearly 9 percent of gross domestic product (GDP). Even when the net economic contribution of the working child is small, it may be important to the welfare of a poor family.

Schooling is typically viewed as the most important alternative use of a child's time outside of work. The net return to school will depend on how the family values future returns to schooling against the direct costs of schooling. Schooling is not the only alternative use of a child's time outside of work. Leisure and play are important components of how children spend their time and may be critically important for child development. In fact, the early Progressive-era arguments about child labor all focused on the value of leisure and play as reasons why children should not work.

Overall, children are most likely to work when the family's valuation of their net economic contribution is high or the perceived returns on alternative uses of the child's time are low. Empirically, poverty has stood out as a key factor influencing the allocation of children's time. Across countries, roughly three-fourths of differences in the economic activity rates of children can be explained by differences in GDP per capita. Within a country, some of the most compelling evidence is from Vietnam, which cut child labor nearly in half over a five-year period during its

economic boom in the 1990s. A majority of this decline in child labor can be explained by improvements in living standards alone.

It is not clear whether the strong poverty child labor connection that is generally observed reflects something about parental or child preferences, changes in the structure of household production, the weakening of credit constraints, a decline in insurance failures, or increases in the returns to activities outside of work such as schooling. All may play some role, and the importance of these factors should vary depending on country context.

Child Labor and International Trade One frequently hears anecdotes about children working in export industries. It is possible that if high-income countries increase labor demand for unskilled labor in low-income countries through international trade, then trade may increase the net economic contribution available to working children and thereby increase child labor. Also, trade can affect returns to education, prices of consumption goods, opportunities for household specialization, and the availability of substitutes for goods or activities involving children.

The cross-country data suggest that the most important connection between trade and child labor stems from the positive association between trade and family incomes. Child labor is lower in countries that trade more because incomes are higher in those countries. This income-driven positive association between trade and child labor holds when one considers all countries, only low-income countries, only trade between high- and low-income countries, and exports of unskilled labor intensive products from low-income countries (Edmonds and Pavcnik 2006). The cross-country data provide no support for the claim that trade perpetuates high levels of child labor in poor countries.

The microeconomic evidence also emphasizes the importance of the effect of trade on family incomes. When income effects are negligible or transitory, children may work more, as apparently occurred in response to coffee price booms in Nicaragua and Brazil (Kruger 2007). When the income effects are positive and long-lived, children work less even when labor demand has increased, as apparently happened with Vietnam's liberalization of its rice trade (Edmonds and Pavcnik 2005). Thus the microeconomic evidence illustrates both that child labor is primarily a facet of poverty and that ultimately decisions about child labor depend on the family's assessment of the relative value of the child's time in its alternative uses.

The End of Child Labor? There is no universally accepted definition of child labor. There is no consensus about how many of the world's 191 million working children are worse off because of their work. Evaluating whether work harms a child requires understanding what children would do in the absence of work. High-quality schooling is too rarely available as an alternative. Hence, it is not surprising that so many of the world's poor families choose to have their children help their families meet their basic needs. Most child work is not in the formal wage sector. In practice, international trade seems to have little influence on the propensity of children to work aside from trade's impact on living standards in low-income countries. Children will no longer work in today's poor countries when families can say that the returns to the child's time in other activities such as school or play are greater than the family's valuation of the child's potential economic contribution.

See also International Labor Organization; labor standards

FURTHER READING

Baland, Jean Marie, and James A. Robinson. 2000. "Is Child Labor Inefficient?" *Journal of Political Economy* 108 (4): 663 79. After Basu and Van, probably the most frequently referenced theoretical paper in the modern literature on child labor. It emphasizes the potential importance of constraints on credit and intergenerational transfers in the parents' decision to have the child work.

Basu, Kaushik, and Pham Van. 1998. "The Economics of Child Labor." *American Economic Review* 88 (3): 412 27. The seminal paper in the child labor literature, framing the persistence of child labor as a coordination problem that could be solved by policy action against child labor.

Edmonds, Eric. 2007. "Child Labor." Working Paper No. 12929. Cambridge, MA: National Bureau of Economic

Research. A recent, comprehensive review of the empirical child labor literature.

Edmonds, Eric, and Nina Pavcnik. 2005. "The Effect of Trade Liberalization on Child Labor." *Journal of International Economics* 65 (2): 401 19. A study of Vietnam's liberalization of rice trade.

———. 2006. "International Trade and Child Labor: Cross country Evidence." *Journal of International Economics* 68 (1): 115 40. Examines the cross country data on trade and child labor in detail.

International Labor Organization. 2002. *Every Child Counts: New Global Estimates on Child Labor*. Geneva: ILO. Reports on the persistence of child labor into the 21st century.

———. 2006. *The End of Child Labor: Within Reach*. Geneva: ILO. The most recent estimates on the number of working children.

Kruger, Diane. 2007. "Coffee Production Effects on Child Labor and Schooling in Rural Brazil." *Journal of Development Economics* 82 (2): 448 63. About the effects of coffee prices on child labor and schooling.

ERIC V. EDMONDS

■ **clusters**

See New Economic Geography

■ **commodity chains**

Commodity chain refers to the linked set of processes involved in the design, production, distribution, and consumption of goods and services in the world economy. Many commodity chains are, and have long been, geographically extensive, spanning multiple countries and regions of the world. They are also often organizationally complex, involving multiple firms and economic agents connected to one another by a range of contractual and noncontractual relationships, including ownership ties, joint ventures, subcontracting networks, and strategic alliances. Because of the analytical leverage that the commodity chain construct provides for understanding the organization of global industries and the dynamics of the capitalist world economy, it has attracted considerable interest among social scientists from a range of disciplines. There are several varieties of global chain analysis, but each focuses on the networks linking people, places, and processes to one another across space, and seeks to underscore the importance of these networks as a critical infrastructure of economic globalization.

History of the Concept The term *commodity chain*, first defined by Terrence Hopkins and Immanuel Wallerstein as a "network of labor and production processes whose end result is a finished commodity," originates in world-systems theory. Three features characterize the world-systems tradition of commodity chain research. First, the focus is on how the global division and integration of labor into the world economy has evolved over time. Although some scholars argue that globalization is a relatively novel process facilitated by advances in information technology and transportation, historical reconstruction of commodity chains suggests that trade and production networks have been international in scope since modern capitalism's emergence. Second, the commodity chain concept enables analysis of the distribution of wealth and power in the world system by focusing on the differential returns to various actors linked through particular chains. Some links in a chain tend to be located in core (i.e., developed) countries of the world system, and others in the less-developed zones of the semiperiphery and periphery, although the particular geography of any chain changes over time. Third, the spatial and social configurations of chains are linked to cyclical shifts in the world economy. World-systems theorists contend that during phases of economic contraction, chains tend to shrink in size (as production volumes or the numbers of producers decline) and/or scope (with production becoming more geographically concentrated). During such periods, the degree of vertical integration along the chain also increases (i.e., a greater number of links become consolidated into fewer). The reverse is true for expansionary periods.

The first book in this field, *Commodity Chains and Global Capitalism*, edited by Gary Gereffi and

Miguel Korzeniewicz, appeared in 1994. This volume featured papers presented at an annual conference of the Political Economy of the World-System section of the American Sociological Association. Most of the essays, with the exception of those on the shipbuilding and wheat flour commodity chains during the 16th and 17th centuries, focus on contemporary manufacturing industries and, in particular, on trade and production networks linking developing country exporters to world markets.

In retrospect, the 1994 volume can be seen as having inaugurated a distinct approach to chain analysis, the global commodity chain (GCC) framework, which diverges somewhat from the world-systems research program on commodity chains. The GCC framework was first developed by sociologist Gereffi, who identified four dimensions with respect to which all commodity chains can be analyzed: (1) an input-output structure, describing the process of transforming raw materials and other inputs into final products; (2) a territoriality, or geographical configuration; (3) a governance structure, referring to the processes by which particular players in the chain exert control over others and appropriate and/or distribute the value that is created along the chain; and (4) institutional context, or the "rules of the game" bearing on the organization and operation of the chain.

Over the course of the 1990s, a substantial empirical literature on commodity chains accumulated, featuring studies of products such as cars, computers, clothing, chocolate, and coffee, among others. By the end of that decade, some scholars were beginning to reappraise the original GCC approach. Specifically, they questioned the very description of these networks as *commodity* chains, since the term commodity is generally taken to denote either primary products (e.g., agricultural staples) or basic manufactures (e.g., T-shirts as "commodity" garments). Others criticized the insularity of the GCC approach, noting that there was relatively little exchange between researchers working within Gereffi's paradigm and those who, although similarly interested in the organizational dynamics of the world economy, were using different concepts to describe interfirm networks in global industries. Some argued that a common terminology would foster dialogue and promote a sense of intellectual community among scholars of international trade and production networks. With the aim of selecting a neutral term that would encompass these various network constructs, *global value chain* (GVC) was chosen "because it was perceived as being the most inclusive of the full range of possible chain activities and end products" (Gereffi et al. 2001).

Governance of Global Chains As Gereffi explained in his contribution to *Commodity Chains and Global Capitalism*, the governance structure of a commodity chain is the set of "authority and power relationships that determine how financial, material, and human resources are allocated and flow within the chain." Gereffi proceeded to make what has become a seminal distinction in the GCC literature between "producer-driven commodity chains" (PDCCs) in heavy manufacturing or more capital-intensive industries such as motor vehicles, and "buyer-driven commodity chains" (BDCCs) in light manufacturing industries such as footwear and apparel. While producer-driven industries tend to be characterized by hierarchy (i.e., links in the chain are vertically integrated within the ownership structure of a firm or, in the case of international production, foreign direct investment), network (i.e., nonequity) forms of governance are characteristic of buyer-driven chains.

Gereffi's BDCC construct was a key theoretical innovation because it pointed to the changing role of commercial capital in establishing and managing global production networks. The lead firms of BDCCs, mostly retailers and brand name marketers, are able to exert control over firms involved in their production networks, although they generally have no equity relationship with the manufacturers and/or contractors making goods on their behalf. One of Gereffi's main interests was to show that even chains with more "marketlike" governance structures require coordination, and that these coordinating tasks are assumed by lead firms that determine much of the division of labor along the chain and define the terms on which actors gain access to it. In this sense, the

distinction between PDCCs and BDCCs presented a new twist on the "markets versus hierarchies" formulation elaborated by Oliver Williamson and other contributors to the new institutional economics: lead firms in BDCCs are frequently connected to their suppliers by networks that differ from both arms-length, one-spot transactions (market) and ownership ties (hierarchy). Nevertheless, these networks are generally characterized by a power structure that gives the "buyers" in these chains leverage over other actors.

Although the analytical utility of these ideal types was confirmed by many studies using the PDCC/BDCC constructs as templates for analyzing various industries, the buyer-driven/producer-driven distinction was also faulted for being too narrow or overly abstract. Some critics suggested that these categories did not adequately capture the range of governance forms observed in actual chains. In his analysis of the electronics industry and the relationships between brand-name computer companies such as Dell and Compaq (so-called original equipment manufacturers, or OEMs) and their major component suppliers, Timothy Sturgeon identified a new governance structure, which he termed the *modular network*. Although modular value chains appear similar to BDCCs in the sense that OEMs typically have no equity tie to the companies manufacturing their components, Sturgeon argued that interfirm networks in the electronics industry are also different from the paradigmatic BDCCs characterizing industries such as apparel.

Sturgeon drew on transaction cost economics in elaborating his theory of value chain modularity. The question underlying transaction cost economics is, why are so many economic activities bundled within the firm instead of transacted on the market? For Williamson, the answer hinged largely on asset specificity; transactions are more likely to be internalized in the firm when they require particular, dedicated investments. This is because such investments foster mutual dependence between the actors in an exchange (for example, between buyer and supplier), which creates conditions for opportunistic behavior on the part of one or both parties to the transaction. To mitigate this risk, safeguards that are capable of guarding against malfeasance must be built into the exchange, and this process increases the cost of the transaction. One way to deal with this problem is through hierarchy that is, the acquisition and incorporation of asset-specific suppliers into the boundaries of a vertically integrated firm. But Sturgeon argues that modular value chains represent a different solution to the problem of transaction costs: in these networks, asset specificity remains relatively low because the codification of knowledge in industry standards allows a highly formalized link at the interfirm nexus between OEMs and their main suppliers.

Thus characteristics specific to the electronics industry particularly the development of industry-wide standards permitting a high degree of codification enable lead firms and highly competent "turnkey" suppliers to exchange rich information (such as detailed product or design specifications) without the kind of intense, face-to-face forms of communication associated with the interfirm networks documented by contributors to the "new economic sociology" literature. Economic sociologists, many influenced by Mark Granovetter's seminal statement regarding the embeddedness of economic activity in social life (1985), have tended to emphasize the "relational" feature of interfirm networks that is, their particular value orientation as open-ended and trust-based. Sturgeon's work implicitly contrasts the relational networks described by sociologists such as Granovetter and Uzzi (1977) with the modular networks found in the electronics industry.

Sturgeon's work on value chain modularity challenged the adequacy of the PDCC/BDCC distinction and underscored the need for a more differentiated understanding of governance structures in global chains. This challenge was taken up by Gereffi, Humphrey, and Sturgeon (2005), who developed a formal theory of global value chain governance that aims to explain and predict the way that exchanges at the interfirm boundary are coordinated. The authors propose a continuum of five governance structures that describe the type of transactional linkage connecting firms in a global

value chain. In addition to the poles of market and hierarchy, this continuum contains three distinct types of network linkages: (1) closest to the hierarchy pole is the captive network, which is characterized by a relatively large power differential between the stronger firm (usually the buyer) and the weaker firm (typically the supplier); (2) the relational network, which is in the middle of the continuum and characterized by complex interactions and greater mutual dependence; and (3) the modular network elaborated in Sturgeon's work on the electronics sector, which is closer to the market pole. Gereffi, Humphrey, and Sturgeon identify three independent variables which, they contend, explain a significant portion of the variation in governance structures across chains: the complexity of transactions, the codifiability of information to be exchanged, and the capabilities of the supply base.

Many GCC and GVC scholars focus on the governance dimension because they seek to understand how the dynamics of global chains can be leveraged into development outcomes, such as facilitating the shift of firms to more profitable links in the chain and fostering the creation of skills and competencies that would permit local producers to upgrade to higher-value-added activities within a chain (e.g., moving from assembly subcontracting to integrated manufacturing), or to switch to a new value chain (e.g., moving from apparel to electronics). For this reason, research on global chains has been supported by bodies such as the UN Commission for Latin America and the Caribbean as well as the U.S. Agency for International Development. The influence of the GCC and GVC literatures in the contemporary development field reflects an elective affinity between global chain frameworks, which analyze the way in which particular economic actors are inserted into international trade and production networks, and the paradigm of export-oriented development, which similarly focuses on the incorporation of local firms and workers into global markets. Chain-inspired development research also highlights, if not resolves, perennial units of analysis problems in development theory and policy: To what extent does increasing the participation of firms in global markets and their competitiveness in particular chains benefit local capital and labor? What is the relationship between firm-level upgrading at the micro level and the more macrolevel development of the regional or national economy?

Future Research and Complementary Approaches As analytical frameworks, the GCC and GVC approaches bring to the foreground the organizational dynamics of contemporary capitalism and their implications for local development in today's global economy. Some critics contend, however, that empirical analysis of commodity chains must pay greater attention to the historical and socio-institutional specificity of the contexts in which these networks are formed and operate. Although ideal types such as the producer-driven and buyer-driven constructs are useful abstractions, the study of commodity chains in situ reveals significant variation in the way that trade and production are organized across space and time, even within the same industry, thus reflecting the influence of the broader regulatory and political-economic frameworks which shape the linkages that emerge between places and processes in the world economy. The contingent and variable nature of international trade and production networks is underscored by recent contributions to an emerging subfield of historical research on commodity chains, with scholars describing and analyzing the centuries-old connections between geographically distant producers and consumers that world-systems theorists coined the term *commodity chain* to describe (Topik, Zephyr, and Marichal 2005).

In addition to the commodity chain concept, several related terms and frameworks have also been developed to analyze the geographical dispersion and organizational fragmentation of production in the global economy. These alternative but potentially complementary frameworks, such as global production networks, systems of provision, and the *filière* approach, draw inspiration from intellectual traditions that differ, to a greater or lesser extent, from those orienting GCC and GVC analysis. Furthermore, the GCC and GVC approaches are not identical to each other, as the former draws more from world-systems theory, organizational sociol-

ogy, and the comparative political economy of development, while the latter borrows from the fields of institutional economics and industrial organization. Despite the different methodological and theoretical emphases of these various approaches, greater dialogue among them may facilitate a richer understanding of how international trade and production networks are organized, as well as their implications for a range of economic and developmental outcomes at the global-local nexus. Similarly, commodity chain analysis might be strengthened by dialogue with the New Economic Geography, since the latter's emphasis on spatial and institutional specificity could enrich the former's understanding of the processes by which the organizational dynamics of global industries become manifest in particular locations.

See also foreign direct investment (FDI); fragmentation; globalization; New Economic Geography; outsourcing/offshoring

FURTHER READING

Arrighi, G., and J. Drangel. 1986. "The Stratification of the World Economy." *Review* 10 (1): 9 74. Applies the commodity chain concept to analysis of inequality and stratification among national economies in the context of world systems theory.

Bair, J., ed. 2008. *Frontiers of Commodity Chain Research*. Palo Alto, CA: Stanford University Press. Collection containing chapters by several key figures including Gereffi, Sturgeon, and Wallerstein, as well as a comprehensive introduction by the editor that provides a broad overview of different approaches to commodity chain research and identifies key methodological and theoretical debates in the literature.

Bair, J., and G. Gereffi. 2001. "Local Clusters in Global Chains: The Causes and Consequences of Export Dynamism in Torreon's Blue Jeans Industry." *World Development* 29 (11): 1885 1903. Explains and evaluates the local implications of an export boom in a region of northern Mexico. Also discusses possible complementarities between the commodity chain and industrial districts literatures.

Gereffi, G. 1994. "The Organization of Buyer Driven Global Commodity Chains: How U.S. Retailers Shape Overseas Production Networks." In *Commodity Chains and Global Capitalism*, edited by G. Gereffi and M. Korzeniewicz. Westport, CT: Praeger, 95 122. Seminal essay introducing the producer driven/buyer driven distinction.

Gereffi, G., J. Humphrey, R. Kaplinsky, and T. Sturgeon. 2001. "Introduction: Globalisation, value chains, and development." *IDS Bulletin* 32 (3): 1 8. Brief introduction to a special issue that introduces the global *value* chain terminology and explains the rationale for its adoption in place of global *commodity* chains.

Gereffi, G., J. Humphrey, and T. Sturgeon. 2005. "The Governance of Global Value Chains." *Review of International Political Economy* 12 (1): 78 104. Elaborates a theory of global value chain governance, including three types of network linkages.

Granovetter, Mark. 1985. "Economic Action and Social Structure: The Problem of Embeddedness." *American Journal of Sociology* 91 (3): 481 510. Widely regarded as a key contribution to the new economic sociology, argues that economic activity is embedded in networks of ongoing social relations.

Henderson, J., P. Dicken, M. Hess, N. Coe, and H. Wai Chung Yeung. 2002. "Global Production Networks and the Analysis of Economic Development." *Review of International Political Economy* 9 (3): 436 64. Provides critical summary of the GCC approach and proposes alternative framework for analyzing dynamics and developmental implications of economic globalization.

Hopkins, T., and I. Wallerstein. 1977. "Patterns of Development of the Modern World System." *Review* 1 (2): 11 145. Introduces the concept of commodity chain in the context of defining a research agenda for world systems theory.

Lane, Christel, and Jocelyn Probert. 2006. "Domestic Capabilities and Global Production Networks in the Clothing Industry: A Comparison of German and UK Firms' Strategies." *Socio Economic Review* 4 (1): 35 67. A comparative analysis of the apparel industry, combining global chain analysis with an institutionalist perspective on national capitalisms.

Ponte, S., and P. Gibbon. 2005. "Quality Standards, Conventions, and the Governance of Global Value Chains." *Economy and Society* 34 (1): 1 31. Insightful critique of GVC analysis, drawing on convention theory.

Sturgeon, T. 2002. "Modular Production Networks: A New American Model of Industrial Organization." *Industrial and Corporate Change* 11: 451 96. Analysis of turnkey suppliers and modular networks in the electronics industry; important precursor to the 2005 Gereffi, Humphrey, and Sturgeon article proposing a theory of GVC governance.

Talbot, J. 2004. *Grounds for Agreement: The Political Economy of the Coffee Commodity Chain.* New York: Rowman and Littlefield. Excellent comprehensive analysis of the coffee commodity chain, drawing on the world systems and GCC traditions of chain research.

Topik, S., F. Zephyr, and C. Marichal, eds. 2005. *From Silver to Cocaine: Latin American Commodity Chains and the Building of the World Economy, 1500 2000.* Durham, NC: Duke University Press. This edited volume features studies of commodities that were Latin America's leading exports during various historical periods, including coffee, bananas, henequen, and rubber.

Uzzi, B. 1997. "Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness." *Administrative Science Quarterly* 42: 35 67. Representative of the "new economic sociology" approach to interfirm networks, emphasizing their embedded and relational nature.

Wallerstein, I. 2000. "Introduction to Special Issue on Commodity Chains in the World Economy, 1590 to 1790." *Review* 23 (1): 1 13. Updating 1986 special issue of *Review* (10, 1) on the same theme, contains analyses of the wheat flour and shipping commodity chains between 1590 1790.

JENNIFER BAIR

commodity-price pegging

Pegging the price of a broadly defined commodity basket would be an almost perfect means to achieve price stability. An adjustable commodity-based dollar was formally laid out in 1920 by the American economist Irving Fisher, who suggested adjustments in the number of resource units in the dollar to achieve price stability. If prices fell below a specific target by 1 percent, for example, the dollar value of the resource unit would be raised by 1 percent   at the same time automatically lowering the number of resource units in the dollar. But the larger the number of commodities involved, the more difficult it is to assign their relative weights. A single commodity price target would be more feasible than Fisher's system; however, it would be desirable only if this single commodity price were well correlated with the aggregate price level.

The principle that governments should intervene in commodity markets to maintain supplies and stabilize prices was endorsed by Confucius in China during the sixth century B.C., and stabilization of grain prices became an important part of early imperial Chinese policy. Communist anti-inflation policy during and after the Chinese Civil War also involved the buying and selling of surplus stocks. In this case, the state established trading companies to undertake large-scale commodity sales in the cities and bring prices down. Later, these same state trading companies used commodity purchases to offset deflationary pressures in 1950 and 1952. On a smaller scale, President Clinton employed the U.S. Strategic Petroleum Reserve (SPR) in a similar fashion in 1996, ordering the sale of a portion of these reserves in an attempt to combat rapidly rising gasoline prices. The SPR was subsequently tapped for limited interventions by the George W. Bush administration as well.

More recently, the economist Jeffrey Frankel suggested that policymakers in smaller commodity-dependent nations use a more dominant form of commodity-price pegging: specifically, policymakers should undertake monetary expansion whenever they face declining demand for that nation's key commodity (Frankel 2003). Under a pure commodity peg, oil producers would simply tie their monetary policy to the price of oil, expanding when oil prices fall and contracting when oil prices rise. Such a strategy has strong intuitive appeal given that it is almost inevitable that an oil-dependent country will face deflationary pressures if oil prices plunge and inflationary pressures if oil prices surge upward. Less specialized producers might enjoy analogous benefits by fixing the price of a basket of export commodities in terms of local currency. Once a target price range has been set, open market operations involving purchases or sales of foreign exchange or domestic

securities could be used to keep the export price index within the band (Frankel 2005). Even larger economies could gain from a commodity-based strategy, at least in comparison to the more popular alternative of pegging to the U.S. dollar. For example, had Argentina adopted a wheat-peg rather than a dollar-peg during 1991 2001, the Argentine peso would have depreciated instead of appreciating as it did over the latter part of the period (Frankel 2003).

The dollar peg was deflationary for Argentina just as the gold peg was deflationary for the United States and many other countries in the late 19th century and during the Great Depression and the authorities in each case ended up maintaining a fixed single-asset price while almost all other prices declined. Faced with very low interest rates and persistent deflation on a global scale during the Great Depression of the 1930s, a number of economists laid out proposals for a "commodity reserve standard" and intervention that would pull commodity prices up from their depressed levels. Some proponents saw this as an alternative to the old gold standard but others saw intervention in commodity markets as simply a means of achieving the desired expansionary end as in, for example, the U.S. silver purchase program that began in 1934 under President Roosevelt.

Other Commodity-Based Proposals Although economists have argued that government warehousing of the commodities involved would not be necessary (see, e.g., Hall 1982), commodity-based proposals have more often been based on government stocks and direct government intervention in commodity markets. For example, one prominent proposal was for a commodity reserve currency to be established. Under this scheme, Federal Reserve banks would buy warehouse receipts for the chosen commodity units so as to "redeem their liabilities, on demand, in commodity units or gold at the option of the holder" (Graham 1941). The premise that a commodity reserve currency could be adopted as an alternative to the old gold standard received serious consideration in the 1930s and 1940s. Critics have pointed to the impracticality and cost of the proposed commodity-market interventions; the relatively narrow range of suitable standardized com-

modities, as well as high elasticity of supply, could force excessively large fluctuations in the stock of money under such a scheme (Friedman 1951).

Although few economists today would advocate an actual commodity-based monetary standard, commodity-price pegging may nonetheless merit consideration as a possible guide for monetary policy. Targeting commodity prices could provide for quantitative monetary easing in the face of deflation, for example, whereas pure interest-rate pegging soon runs out of room as interest rates cannot be driven below zero. In this case, implementation of commodity-price pegging could help secure an inflation or price-level target or, at the very least, offer a complementary policy goal.

See also currency board arrangement (CBA); dollar standard; exchange rate regimes; Federal Reserve Board; gold standard, international; impossible trinity; money supply

FURTHER READING

Burdekin, Richard C. K., and Fang Wang. 1999. "A Novel End to the Big Inflation in China in 1950." *Economics of Planning* 32 (3): 211 29. Documents the commodity market intervention that played a key role in Communist stabilization policy in China in the late 1940s and early 1950s.

Fisher, Irving. 1920. *Stabilizing the Dollar*. New York: Macmillan. Includes Fisher's original proposal for a commodity based dollar.

Frankel, Jeffrey. 2003. "A Proposed Monetary Regime for Small Commodity Exporters: Peg the Export Price." *International Finance* 6 (1): 61 88. Makes the case for smaller resource based economies fixing the local currency price of their chief commodity export.

. 2005. "Peg the Export Price Index: A Proposed Monetary Regime for Small Countries." *Journal of Policy Modeling* 27 (4): 495 508. Generalizes Frankel's earlier proposal to provide for fixing the price of a broader basket of exports for countries with a more diversified pattern of production.

Friedman, Milton. 1951. "Commodity Reserve Currency." *Journal of Political Economy* 59 (3): 203 32. Questions the feasibility of introducing a commodity based currency standard and considers the large fluctuations in the stock of money that might result.

Graham, Frank D. 1941. "Transition to a Commodity Reserve Currency." *American Economic Review* 31 (3): 520 25. A representative proposal for a commodity based currency backed by warehouse receipts for the selected commodity units.

Hall, Robert E. 1982. "Explorations in the Gold Standard and Related Policies for Stabilizing the Dollar." In *In flation: Causes and Effects*, edited by Robert E. Hall. Chicago: University of Chicago Press, 111 22. Shows that a particular mix of ammonium nitrate, copper, aluminum, and plywood would have closely tracked the U.S. cost of living over the 1946 80 period.

RICHARD C. K. BURDEKIN

■ Common Agricultural Policy

The Common Agricultural Policy (CAP) of the European Union (EU) is called "common" because the main agricultural policy decisions are made at the EU level, and these agricultural programs are financed for the most part from the EU's common budget. The functioning of the agricultural sectors in the member countries of the EU is widely determined by decisions at the Community level. Consequently, agriculture is perceived as one of the most integrated sectors in the EU and is sometimes viewed as a possible model for other sectors. Although the EU's agriculture sectors are politically integrated, however, they are less integrated economically than other sectors.

The creation of the CAP in 1962 was vital for the European Economic Community (EEC), the forerunner of the EU. The six founding countries Belgium, France, Germany, Luxembourg, the Netherlands, and Italy had their own national agricultural policies, which differed vastly. The individual countries were reluctant to remove their intervention in agricultural markets, or allow for free trade among member countries and to follow the rules of a customs union. It did not make sense to exclude one sector from market integration, however, and politically the stakes were high: France wanted access to the large German market for its agricultural products if it were to allow for German

industrial products to enter the French market. Hence, it became necessary to establish the organizational and institutional framework for agricultural policy at the Community level.

The institutional framework set up at the inception of the CAP was a compromise to satisfy the heterogeneous interests of the member countries. Under German pressure, the EEC decided on a common price level for agricultural products significantly higher than world market levels and introduced border regulations that completely disconnected EEC prices from those on the world market. The evolution of the CAP has proven that this initial framework constrained the CAP in adapting to changing conditions in the EEC and on the world market.

The Features of the CAP at the Time of Inception Common market organizations regulated the CAP for the main agricultural products. These market organizations are Community law and include the legal framework needed to administer the markets. The aim was to keep domestic farm prices above the level of world market prices. Variable levies were implemented as border measures to cover the gap between the domestic farm prices and the corresponding world market prices. The levies varied with the world price since the commonly agreed-upon prices were the lowest prices at which foreign supply was allowed to enter the European domestic market. From the start, the CAP provided for export subsidies in the event that domestic production was to surpass domestic use at domestic prices. The export subsidies were termed "export restitutions." Under the General Agreement on Tariffs and Trade (GATT) and later the World Trade Organization (WTO), export subsidies for industrial products were eliminated and import tariffs rates were bound by a ceiling rate negotiated by member countries. Thus the trading regime for agricultural products remained very different from that of industrial products.

In addition, market organizations included domestic measures of protection. The most important ones were guaranteed purchases by state agencies of selected products such as butter, skimmed milk

powder, grains, and sugar at intervention prices, and facultative purchases for some other products if some additional conditions prevailed. The protection regime for sugar was exceptional as it included intervention prices for sugar, minimum purchase prices for sugar beets, and production quotas. For quantities above their allocated quotas, farmers received the world market price. The sugar market regime was the least in line with free market principles; it was more comparable to instruments used in centrally planned economies.

The Evolution of the CAP The institutional prices were the main instruments of the market regimes that constituted the Community law. The Community law or any changes to the law were generally proposed by the European Commission; in the case of agricultural law the decisions were made by the Council of Ministers, which is composed of agricultural ministers from member countries. The most important decision by the Council of Ministers was to make the agreed changes to the institutional prices. These changes were made annually up to the 1990s. Since no objective criteria guided the setting or changing of institutional prices, political considerations dominated these decisions. National interests diverged widely between member countries, partly because of differences in their national conditions such as inflation, overall growth in the economy, and productivity changes in agriculture. Moreover, the framework that was instituted to finance these policies accentuated the differences in national interests among the member countries. The EU had adapted the principle of financial solidarity, which stipulated that policies implemented at the Community level should be financed from a common budget. As the EU did not generate revenues, the contribution to the common budget came either from the member countries' budgets or from import tariff revenues, that is, member countries would deliberately forgo their own tariff revenue in favor of the Community budget. Not all countries benefited from these policies. For example, the Netherlands was traditionally a huge exporter of dairy products so decisions that resulted in higher dairy prices increased the income of Dutch dairy farmers much

more than it taxed its consumers. For the United Kingdom (UK), which was a huge importer of dairy products, an increase in dairy prices had the reverse effect. Hence, it is understandable that national interests diverged significantly.

The voting procedure made it even more difficult for member countries to come to an agreement. The Treaty of Rome (1957), from which the Community emerged, provided for majority voting for most decisions. The six founding members agreed in 1966, before the first market organization came into existence in 1967/68, to apply unanimity voting whenever there were vital interests at stake for any individual country. Thus agreements could be reached only if member countries for which common policies conflicted with national interests were compensated. The result was that an increase in the domestic price level was decided year after year, even in times where world market prices for agricultural products declined and the exportable surplus of the EU rose. The increase in agricultural protection, the enlargement of the EU, and unprecedented technological progress in EU agriculture contributed to a faster growth in EU agricultural production than consumption. For the Community, which had started as an importer of agricultural products, the changes in supply and demand reduced the import gap gradually over time and by the end of the 1970s generated export surpluses. Increasing exports and falling world market prices put pressure on the EU budget as outlays for export restitutions rose strongly. A new phase of the CAP began with even more governmental interference in the market. The milk market presented the most urgent problem. Milk production in the EU grew strongly as imports of concentrated feed could be imported free of duties due to the GATT. At the inception of the CAP, the EU had set its bound tariffs at zero or a very low level at a time when imports of these products were negligible. In return, the EU was allowed to introduce the variable levy system. The import regulation for concentrated feed became the Achilles' heel of the CAP. Low prices for feed, which were partly due to significant declines in transatlantic transport costs, and high support prices for CAP products, led to high

effective rates of protection for meat and milk and increases of production. On the demand side, imported feed was substituted for domestically produced grain, and because of cheap imports of vegetable oil, margarine was substituted for butter. The consequence was the accumulation of stocks in the EU and accelerating budget outlays. The latter seemed to have been the main constraint for the CAP during this period.

In 1984, the council agreed on a quota system for milk. Farmers received a right to sell under the quota at guaranteed prices. Sales above the quota were taxed with a superlevy, which for many years was higher than the price paid for sales within the quota. Hence, production was curtailed and growth of production was stopped. Moreover, a set-aside program was offered to farmers in 1986. Grain farmers were offered a premium if they voluntarily took arable land out of production.

The first 25 years of the CAP were characterized by increases in border and domestic EU protection for agricultural products. The EU reacted to changes in the economic environment mainly by introducing new instruments to limit budgetary expenditure. The concerns of the trading partners were for the most part neglected.

1992 to 2003 The Uruguay Round (1986 94), the last round under the GATT, focused on agricultural trade. The EU had to reform the CAP in order to comply with the rules of the new round. The council decided to first reform the EU grain market in 1992. For the first time, the CAP prices of specific commodities were cut, some significantly: the council agreed to cut institutional prices for grain by 30 percent. Farmers were compensated entirely for their prospective income loss by direct payments. These payments were linked to the use of land for grain. In addition, farmers would qualify for payments if they set aside at least a politically determined percentage of their land. Although the effect on grain production was minimal, domestically produced grain became competitive against imported concentrated feed and, thus, the grain surplus vanished.

The decision on the grain sector contributed to a final agreement in the Uruguay Round. The agreement required further changes to the CAP, however. The EU's export subsidies were now constrained in two ways:

1. the quantity of well-defined products or product groups that was allowed to be subsidized was reduced by 21 percent (from their 1986–88 level); and
2. the amount of export subsidies had to be cut by 36 percent (from their 1986–88 level).

Moreover, like other GATT/WTO members, the EU had to accept concessions on the import side. Each country was to allow imports of specific products up to 5 percent of domestic consumption in the base period 1986 88; variable levies and other nontariff trade measures had to be converted to tariff equivalents (tariffication), and the tariff equivalents had to be reduced by 36 percent on average from the base period 1986 88, and by at least 15 percent for individual products. Thus, the EU was not allowed to apply the variable levy system anymore and had to take into account changes in world prices when setting domestic institutional prices. It was this international agreement that forced the EU to change its market organizations. Finally the EU reformed the CAP drastically in 2003, aiming at making a positive contribution to the Doha Round of trade negotiations, which was launched in November 2001. One of the reforms was of domestic support. The CAP direct payments were included in the "Blue Box." These are direct payments under production-limiting programs and hence they were exempted from reduction. This concession was given to the EU in return for price cuts and the realization that aid to the farmers was needed during the adjustment period. But the EU would have to give up these exemptions by the end of the Doha Round as adjustment compensations cannot be paid permanently, and the EU would have to look for alternatives for providing aid to farmers.

CAP Reform in 2003 and Thereafter The CAP reform of 2003 was a change in the paradigm of agricultural protection. Previously, support to agri-

culture took the form of price support to output and factors of production; and in the livestock sector, some kind of payments were made by head of animals. This support stimulated production and taxed consumption. Hence trading partners opposed this system and pressed for a change. The council decided in 2003 to reduce price support even more (for the main products from 2004 onward and for some products such as sugar, olive oil, tobacco, and cotton at a later time) and to decouple in principle all types of direct payments. The general proposal of the European Commission combined all income losses due to price cuts made until 2003 and all types of direct payments granted up to 2003 in one type of payment, the Single Farm Payment, to be made to individual farms based on historical levels. Entitlements to farm payments were made tradable and, thus, became decoupled completely from production, at least in principle. But the council allowed member countries a lot of flexibility. For example, countries were allowed to link part of the payments to production or to the use of factors of production during the transition period and to link them again to production in the future if regional or domestic production in a country fell too strongly. Payments could also be linked completely to land, either by granting a flat rate across the country or by differentiating between regions or between arable land and grassland. Finally, some payments were tied to environmental standards that farmers needed to meet, a process known as cross-compliance. Cross-compliance links payments to farmers to their respect of environmental and other requirements (such as animal and plant health and animal welfare) set at EU and national levels.

The 2003 reform also introduced a new classification of the CAP instruments: Pillar I for market and price support measures and Pillar II for rural development. In the reforms, a decline was imposed on Pillar I spending, while Pillar II spending was allowed to go up.

The distinction between the two types of measures may sound economically reasonable, but what justification could there be for increasing Pillar II's budget by about the same amount that Pillar I's decreases, given that in principle the two pillars have different purposes? Rural development should not focus mainly on agriculture. Rural development needs differ across regions and countries and they are not related to past agricultural support, so it is questionable that the changes in budget needs for rural development of each region and country (Pillar II) match widely the reduction made in agricultural support (Pillar I).

Hence critics suspect that Pillar II is a way for policymakers to continue providing support to farmers when they can no longer justify direct payments as adjustment compensation for price cuts that occurred more than a decade earlier. The framework of Pillar II seems suitable for hiding subsidies for agriculture. Pillar II measures require a higher quality of governance than Pillar I measures, but unfortunately, due to the enlargement of the EU from 15 to 25 member countries in 2004, and the further enlargement on the inclusion of Romania and Bulgaria in 2007, the competence to monitor and to enforce Pillar II measures has declined on average. The European Court of Auditors regularly finds irregularities and fraud in Pillar II measures.

The Future of the CAP The CAP has changed significantly over time. Changes have been driven by budgetary concerns rather than economic rationale, however. Up to the 1990s, budget constraints led to reforms, but they resulted in even more protective and trade-distorting policies. Finally, under the GATT/WTO multilateral trade negotiations, the CAP abandoned most of its border protection. Some products such as sugar and milk continue to be highly protected, however. Changes in the EU and other countries' policies and rising oil prices may result in higher world prices and may make further reforms of EU border measures unnecessary. But if these changes do not happen, the EU will need to cut its agricultural prices further in order to comply with the constraints imposed by the WTO.

The role of the CAP has become more important for the world food system over time as the EU has grown from 6 to 27 members as of 2007. The

reduction in the CAP's external rate of protection contributes to less distorted agriculture. Two major concerns remain, however: first, as long as EU prices for food products are not completely linked to world market prices, the CAP does not buffer fluctuations in world market prices to the maximum possible extent. Thus further complete coupling of EU and world market prices is needed. Second, Pillar II of the CAP allows for hidden farm subsidies and causes distortions across EU member countries and worldwide. Hence, Pillar II measures have to be reconsidered and scrutinized with respect to distortive effects.

See also agricultural trade negotiations; agriculture; common market; distortions to agricultural incentives; European Union; political economy of trade policy

FURTHER READING

Borrell, B., and L. Hubbard. 2000. *Global Economic Effects of the EU Common Agricultural Policy.* Oxford: Institute of Economic Affairs, 18 26. An identification of the costs caused by CAP based on a quantitative model.

Brümmer, B., and U. Koester. 2004. "EU Enlargement and Governance of the CAP." In *European Union: Challenges and Promises of a New Enlargement*, edited by A. Puskas. New York. International Debate Education Association, 226 46. The focus is on governance problems of the CAP.

European Commission, Directorate General for Economic and Financial Affairs. 1994. "EC Agricultural Policy for the 21st Century." Reports and Studies No. 4. Brussels. An assessment of the evolution of the CAP and a pro posal for change.

Frandsen, S., and A. Walter Jorgenson. 2006. "The Common Agricultural Policy: A Review." In *WTO Negotiations and Agricultural Trade Liberalization: The Effect of Developed Countries' Policies on Developing Countries*, edited by E. Diaz Bonilla, S. Frandsen, and S. Robinson. Wallingford Oxfordshire, UK: CAB, 34 56. A nontechnical survey of the evolution and functioning of the CAP.

Koester, U., and A. El Agraa. 2007. "The Common Agri cultural Policy." In *The European Union, Economics, and Policies*, edited by A. El Agraa. 8th ed. Cambridge University Press, 2007, 373 410. A survey of the

functioning and evolution of the CAP, which can be used as classroom text.

ULRICH KOESTER

■ common currency

In the modern world economy, a common currency shared by a number of countries has been a rare phenomenon. At least since the 19th century, the rule has been that independent countries have independent currencies. Territorial currencies were established in the 1815 1914 period, a result of important technical advances such as presses that could mint coins and print notes that were hard to counterfeit combined with economic factors, such as the spread of the monetized economy and the state taking on more functions and casting its revenue net more widely. As new nations such as Italy, Canada, and Germany formed in the second half of the 19th century, they adopted national currencies.

The pattern of one country one currency was reinforced with the decline of empires. Some former colonies were reluctant to discard shared currencies, but after a brief time lag, currencies became national. An example from the 1990s is the rapid introduction of national currencies following the dissolution of Yugoslavia, the USSR, and Czechoslovakia. The 12 Commonwealth of Independent States successors to the USSR recognized the increased transaction costs that would result from abandoning the common currency, but after less than two years the ruble zone had collapsed.

Most currencies are national, where N_m is the number of monies and N_c is the number of countries:

$$N_m \approx N_c \qquad (1)$$

In practice, however, the one country one currency rule is not always the case and sometimes $N_m < N_c$. The exceptions to the rule fall into two categories: carrots and ministates (Pomfret 2005).

Carrots and Ministates The principal examples of currency union driven by carrots (i.e., incentives offered by an interested party) are the CFA franc zone in central and West Africa and the rand zone in

southern Africa. The French Treasury, which acts as a guarantor of the fixed exchange rate, manages the CFA franc zone's reserves and settles the regional central banks' payments and receipts. The zone has existed for more than a half a century because of preferential French aid to zone members and balance of payments (BOP) support. In the rand zone, South Africa has formal arrangements to share seigniorage (i.e., the real resources obtained from printing money which can be spent on goods and services) with the countries in which the rand is legal tender (Lesotho and Namibia), and the South African central bank is prepared to act as lender of last resort in these countries. Similar carrots encouraged retention of the ruble zone in 1992 93, but Russia objected to the size of transfers to other members while some members opposed the political use of the levers. Ultimately, lack of agreement on monetary policy institutions made the ruble zone unstable. In the both the CFA franc zone and the rand zone, by contrast, members accept the institutions imposed by a dominant economic power. Even so, the membership of the CFA franc zone and the rand zone has not been entirely stable. Mali withdrew from the CFA franc zone in 1962 and rejoined in 1984; Mauritania withdrew in 1973; Equatorial Guinea, a former Spanish colony, joined in 1985; and Guinea-Bissau, a former Portuguese colony, joined in 1997. Botswana withdrew from the rand zone in 1976.

Other countries that lack independent currencies or those that participate in a shared currency arrangement are small and often special cases, known as ministates. Between 1970 and 1990, according to Rose (2000, 41), 82 countries were involved in currency unions, 15 of which were in the CFA zone and three in the rand zone. The remaining 64 countries were small economies using the currency of a neighboring or quasi-colonial power or ceding monetary policy control to a larger country. The Eastern Caribbean Currency Area (consisting of eight small island economies), the British Virgin Islands, Bahamas, Barbados, and Belize have tied their currencies' value to the U.S. dollar since 1976. Ireland-UK (pre-1979), Luxembourg-Belgium (pre-euro), and Brunei-Singapore are often described as

currency unions, but the second-named country in each pair had total control over monetary policy. The remainder of the 82 currency union members used another country's money for all or part of the twenty-year period. Most of these are tiny economies, such as Svalbard, Isle of Man, and Norfolk Island. The largest, Liberia, ceased using the U.S. dollar in the 1980s when a new government started issuing first coins and then paper currency. The next largest, Panama, has had a special status with the United States since it was created prior to construction of the Panama Canal.

Historically, some colonies have had separate currencies, although this typically involved large colonies of a kind that no longer exists. The status of some territories is not clear cut, because the ruler, occupier, or guardian is sensitive to the term *colony*, but territories such as Guam (and Panama, to a lesser extent), Northern Cyprus, or Western Sahara have a quasi-colonial relationship with the country whose currency they use. Other currency union members are integral parts of the larger nation. Scotland, for example, is part of the United Kingdom, not a member of a currency union, just as Christmas Island is part of Australia.

Apart from the CFA countries, Lesotho, Namibia, Swaziland, Ireland, Luxembourg, Brunei, Liberia, and Panama, all currency union members identified from 1970 to 1990 were small islands or territories without full sovereign status. An updated version of this list would include another exception: the eurozone. The euro, however, is a unique example of independent nation-states agreeing to use a common currency whose monetary policy is determined by a common institution.

Explaining Currency Domains The theoretical approach to explaining the use of a common currency has been dominated by the optimum currency area (OCA) literature initiated by the economists Robert Mundell and Ronald McKinnon in the early 1960s and continuing to Alberto Alesina and Robert Barro (2002). OCA theory, which emphasizes the trade-off between the macropolicy benefits of an independent currency and the microeconomic benefits from a common currency, is appealing in

principle but has a poor record of explaining the composition of existing currency areas or predicting changes in currency domains.

Independent countries want their own currency because they want to set their own monetary policy. Economic stability is not always a primary concern. Ukraine left the ruble zone in 1992, for example, so that the government could print money to support inefficient producers. In many newly independent countries in all eras, monetary policy independence enabled rulers to finance expenditures. Fixed exchange rate arrangements and currency unions both involve some level of constraint on independent monetary policy, but all exchange rate regimes involve an important element of choice (over instrument and peg) and leave an option of reversal. Fixed exchange rate systems enable countries to make their own macroeconomic policy choices, while a currency union definitively cedes control over monetary policy.

OCA theory takes the position that very small nations, or microstates, cannot afford to have a national currency because the transaction costs would be too high. The threshold for such transaction costs is not so high as to prohibit small countries like Malta or Iceland from having independent currencies, however. Neither are transaction costs the reason for membership in the CFA franc zone and rand zone; these members use common currencies because France and South Africa provide carrots for them to do so, and even then the carrots were insufficient to keep Mauritania in the franc zone or Botswana in the rand zone.

What explains the use of a common currency by independent countries, and especially the case of the euro? The OCA theory does not explain the timing or composition of the euro because it ignores the importance of a common currency for public finance. For governments, seigniorage is a benefit of a national currency, but the benefit tends to be small, especially as cash declines in significance. More important is the need to have a common unit of account for the public finances; the concept of legal tender allows the government to set tax rates, approve expenditures, and similar functions. The use of multiple currencies within one country would undermine political

agreement over the allocation of the central budget. Internal exchange rate changes would reduce the tax burden and increase the relative value of expenditures for the users of one currency over another.

The Adoption of the Euro The emergence of single-currency areas paralleled the consolidation of the nation-state. The major currency unions of the second half of the nineteenth century, including those in Germany, Italy, and Canada, were associated with political unions. In 1990, German monetary union accompanied the reunification of the country. Although debates took place at the time over acceptance of the deutschmark in the former East Germany and the rate at which old Ostmarks would be exchanged, a common budget in common units played a key role in German reunification. The euro, first adopted for noncash transactions in the late 1990s before it became a cash currency, is a 21st-century example of public finance driving adoption of a common currency.

The adoption of the euro was preceded by a lengthy transition period during the existence of the European Monetary System, which began in 1979, and especially after the Maastricht agreement of the early 1990s. The process did not follow the predictions of the OCA theory. Although the EU did become more integrated with more open national economies and greater movement of labor and capital across national borders, the pace of monetary integration did not follow these trends, and in the end, restrictions on capital movements were abolished as a step toward monetary union rather than monetary union being driven by greater factor (e.g., labor or capital) mobility.

How to explain these outcomes? The European Monetary System began operation in 1979, and an important driving force was the difficulty of managing the Common Agricultural Policy based on agreed common prices when exchange rates were market-driven (Pomfret 1991; Basevi and Grassi 1993). The problems of the EU's agricultural policy may be reduced by reform, but any common policies based on political negotiations about financial contributions and monetary benefits will be undermined by changes in bilateral exchange rates. The

more far-reaching the EU's common policies and the larger the EU's common budget became, the more severe the problems associated with lack of a common currency. Meanwhile, the desire for independent monetary policies was moderated by growing agreement on the primacy of price stability and on the desirability of central bank independence. In the 1990s, the crucial issues behind adoption of a common currency concerned who determines the conduct of monetary and fiscal policy, rather than the emphasis in OCA theory on private sector transaction costs and on whether macropolicy will be effective or not.

If the EU is becoming a territorial unit as Germany or Italy or Canada did in the 19th century, and this was a significant motive behind the introduction of the euro, then the euro as a common currency is sui generis in the current world economy. Apart from the euro, the only current examples of common currencies are the special cases of the CFA franc zone and the rand zone, where carrots encourage membership, and the extreme cases of ministates.

See also Bretton Woods system; dominant currency; euro; European Monetary Union; exchange rate regimes; impossible trinity; multiple currencies; optimum currency area (OCA) theory

FURTHER READING

Alesina, Alberto, and Robert J. Barro. 2002. "Currency Unions." *Quarterly Journal of Economics* 117 (2): 409 36. A survey of the theoretical literature.

Basevi, Giorgio, and Silvia Grassi. 1993. "The Crisis of the European Monetary System and Its Consequences for Agricultural Trade." *Review of Economic Conditions in Italy* 47 (June): 81 104. An analysis of the relationship between the EU's agricultural policy and its decision to push toward monetary union in 1978 79.

Bowles, Paul. 2007. *National Currencies and Globalization: Endangered Specie?* London: Routledge. An analysis of the threat posed by globalization to the national currencies of medium sized countries like Norway or Canada.

Helleiner, Eric. 2003. *The Making of National Money.* Ithaca, NY: Cornell University Press. A historical account of the rise of national currencies over the last two centuries.

Mckinnon, Ronald I. 1963. "Optimum Currency Areas." *American Economic Review* 53 (September): 717 24. One of the pioneering papers on the OCA theory.

Mundell, Robert A. 1961. "A Theory of Optimum Currency Areas." *American Economic Review* 51 (November): 509 17. One of the pioneering papers on the OCA theory.

Pomfret, Richard. 1991. "The Secret of the EMS's Longevity." *Journal of Common Market Studies* 29 (December): 623 33. An analysis of why European monetary integration went off track in the 1970s but not in the 1980s, emphasizing the importance of fiscal considerations.

———. 2005. "Currency Areas in Theory and Practice." *Economic Record* 81 (253): 166 76. A review of the empirical applicability of the main theories of currency area formation.

Rose, Andrew. 2000. "One Money, One Market: The Effect of Common Currencies on Trade." *Economic Policy* 30 (April): 7 33. A seminal and controversial assessment of the impact of a common currency on bilateral trade.

RICHARD POMFRET

■ **common market**

A c*ommon market* differs from other preferential trading arrangements, such as a free trade area or a customs union, in that, in addition to free trade in goods and services among members, there also is free mobility of factors, that is, free mobility of labor, capital, and other inputs used in producing goods and services. The European Union (EU) is a prominent although imperfect example of this form of integration.

Global free trade and global factor mobility allow the most efficient allocation of resources and, therefore, yield the best situation. In contrast, a common market, which restricts trade and factor mobility between members and nonmembers, has to be seen in a *second-best* context. As in any preferential trading arrangement, a common market leads to welfare-reducing trade-diversion effects to go along with its welfare-enhancing trade-creation effects. Trade diversion occurs when an importing country buys a

good at a higher price from a member of the common market, while this same good was previously imported at a lower price from a nonmember nation. This can happen because of the tariff preference that is granted to the members. As a result, trade is diverted from a low-cost nonmember to a high-cost member, and this is efficiency reducing. These effects apply to factor mobility as well. To the extent that intrabloc capital movement or labor migration might improve factor allocation within the union, it is efficiency enhancing. On the other hand, if factor mobility within the union leads a country to replace factor inflows from nonmembers with less-efficient factors from members, it will be welfare reducing. Finally, as with goods, the prices for mobile factors might differ in a common market, thereby conferring terms-of-trade losses or gains to member nations.

In the presence of foreign factors of production, many of the conventional welfare results of international trade break down. Brecher and Bhagwati (1981) highlight the role of the *differential trade volume* phenomenon and the *differential trade pattern* phenomenon in leading to apparent paradoxes. The trade volume for the home nation as a whole will, in general, differ from the corresponding volume for the nationals (excluding the immigrant factors of production) only. This is referred to as the *differential trade volume*. It is also possible that, while the country is a net exporter of (say) a capital-intensive good, the nationals are net importers of the same good. Brecher and Bhagwati (1981) call this the *differential trade pattern*. A terms-of-trade improvement for the country can be an adverse movement for the nationals because of this effect. Similarly, even with the same trade pattern, an adverse terms-of-trade movement will be amplified if the nationals' volume of trade is disproportionately large (compared to the country's as a whole). This can reduce the income of the nationals, even if the aggregate income (including that of the immigrants) rises. Overall, they conclude that even if standard trade theory may suggest that a nation will gain (or lose) from a certain change, the effect on its citizens might be ambiguous in the presence of immigrant factors. In light of this discussion, we infer that factor mo-

bility can lead to conflicts between member-nation interests. Policies that might be good for the union need not raise the national income of each member. In such a situation, welfare-enhancing policies might be feasible only if members can coordinate and set up compensation mechanisms that allow for appropriate intrabloc transfers.

Wooton (1988) considers the effects of moving from a customs union to a common market. To the extent that factor mobility leads to superior resource allocation within the union, efficiency rises. However, given that trade taxes exist between the union and the rest of the world, factor mobility can amplify or reduce distortions. For example, consider a good that is subject to an import tariff by the union. When factors move within the bloc, production of the good will fall in some nations and rise in others through changes in their respective production possibility frontiers. If the net effect is such that production of the good falls at the union level, more will be imported by the union, thereby moving the outcome closer to the free-trade level. In such a situation, a common market will improve on a customs union. If not, then the opposite will occur. The welfare issue is more complicated at the national level: even if the union as a whole gains, some member nations might lose because of adverse terms-of-trade movements in product and/or factor markets. For an individual member nation, the extent of gains or losses from such movements will depend on its volume of trade. Furthermore, because a common external tariff restricts trade, a member nation will gain if factor mobility induces an expansion of trade with nonmembers. Kowalczyk (1993) focuses on related issues and, in addition to the terms-of-trade and volume-of-trade issues related to trade taxes, considers the role of nontariff barriers (NTBs). He points out that intraunion trade in goods or factors is often subject to NTBs that do not generate revenues. Consequently, the welfare gains related to NTB removal depend on the initial volume of trade and not the change in the volume, which is scaled by the tariff rate. Viewed in this context, a union that already has a large volume of trade in factors is likely to gain significantly from the removal of impediments to factor mobility.

Issues of tax competition and coordination also arise when there is factor mobility between union members (see Haufler 2001). Broadly speaking, as long as there are differences in national policies within a union, there will be incentives for labor and/or capital to move to take advantage of the differences. By the same token, nations can anticipate such movements and adjust their policies accordingly. In the absence of coordination at the union level, factor mobility can lead to tax competition between nations and to inefficient policy outcomes for an individual member and perhaps for the union as a whole. Competition for mobile capital is an example: Consider two nations that are competing to have capital locate within their borders. Assume also that capital-tax revenues go toward financing a public good. If one nation raises its tax rate on capital, the other nation will benefit as some of the capital will relocate to it. In other words, the tax imposed by one nation causes a positive externality on the other, which experiences an increase in tax revenue, and public-good provision, without increasing its tax rate. In a noncooperative Nash taxation equilibrium, this leads to tax rates and public-good provision that are too low from the perspective of a union of these two nations. At a noncooperative Nash taxation equilibrium, each country sets its capital tax unilaterally to maximize its objective function, assuming that the other country's tax rate is fixed at a certain level. In addition, at this equilibrium, the tax rates of the two nations (say A and B) must be such that, given A's tax rate, B's tax rate maximizes its objective function, and vice versa. Tax coordination between the nations will help to alleviate this problem. However, asymmetry between nations can make such coordination difficult. It is possible that the coordinated outcome is not superior for a nation that stands to benefit from tax competition. Therefore, for heterogeneous unions, coordination can be harder to achieve.

Although capital and labor mobility have been treated symmetrically in our discussion so far, there is evidence that actual movements of labor within a common market are relatively small compared to the movement of capital. Dustmann et al. (2003) note that the introduction of mobility between Greece, Portugal, and Spain and existing EU members did not lead to large migration flows from these nations to the rest of the EU. This is significant, because these acceding countries differed substantially in their economic conditions (for example, much lower per capita incomes) relative to the existing EU nations. Qualitatively similar conclusions are reached by Zaiceva (2004) regarding accession of East European nations to the EU. No major jumps were anticipated and, indeed, as income levels converge over the long run, migration flows are expected to fall. As an explanation for the low levels of labor migration following integration, Dustmann et al. suggest that it is not only current conditions but also expectation of future conditions that can determine such flows. To the extent that agreements such as that for EU expansion create optimism in a new member nation about its own economy, it may dampen the desire of a potential migrant to incur the costs associated with migration.

Harris and Schmitt (2005) contrast intra-EU mobility with the relatively high mobility between states in the United States and note that labor market shocks in the EU generate changes in the labor market participation rate without affecting migration significantly. The EU experience may not immediately generalize to other groupings of nations. For example, mobility between Canada and the United States has been high historically and is likely to strengthen with potentially deeper labor market integration under the North American Free Trade Agreement (NAFTA) (see Harris and Schmitt 2005). Similarly, labor market integration that includes Mexico might also lead to large flows. The migration pressure between Mexico and the United States is readily seen in the large existing stock (and flow) of legal and illegal immigration across the U.S.-Mexican border. At least in the context of NAFTA, however, the recent immigration reform proposals, and the debate surrounding them, suggest that a common market with full mobility of labor seems unlikely to be politically feasible in the near future.

See also customs unions; European Union; free trade area; North American Free Trade Agreement (NAFTA); regionalism

FURTHER READING

Brecher, Richard A., and Jagdish N. Bhagwati. 1981. "Foreign Ownership and the Theory of Trade and Welfare." *Journal of Political Economy* 89 (3): 497 511. A theoretical contribution qualifying standard trade theoretic results in an economy with immigrant factors of production.

Dustmann, Christian, Maria Casanova, Michael Fertig, Ian Preston, and Christoph M. Schmidt. 2003. "The Impact of EU Enlargement on Migration Flows." Home Office Online Report 25/03. Center for Research and Analysis of Migration (CReAM), UK. Assesses potential migra tion flows to the UK due to EU enlargement.

Harris, Richard G., and Nicolas Schmitt. 2005. "Labour Mobility and a North American Common Market: Implications for Canada." In *Brains on the Move: Essays on Human Capital Mobility in a Globalizing World*, edited by S. Easton, R. G. Harris and N. Schmitt. Toronto: C. D. Howe Institute, 133 74. Focuses on greater labor mobility within NAFTA and its implica tions for Canada.

Haufler, Andreas. 2001. *Taxation in a Global Economy.* Cambridge: Cambridge University Press. Brings to gether different policy coordination issues that may arise within an economic bloc like the EU.

Kowalczyk, Carsten. 1993. "Integration in Goods and Factors: The Role of Flows and Revenue." *Regional Science and Urban Economics* 23: 355 67. Compares revenue generating barriers to trade to NTBs, in the context of a common market type of trading bloc.

Wooton, Ian. 1988. "Towards a Common Market: Factor Mobility in a Customs Union." *Canadian Journal of Economics* 21 (3): 525 38. Assesses efficiency gains for members of a customs union, who plan to move to a common market type of integration.

Zaiceva, Anzelika. 2004. "Implications of EU Accession for International Migration: An Assessment of Potential Migration Pressure." CESifo Working Paper #1184. Empirical forecast of migration flows from accession nations to EU member states due to eastern enlarge ment.

SUBHAYU BANDYOPADHYAY
AND HOWARD J. WALL

■ **Common Market for Eastern and Southern Africa (COMESA)**

The Common Market for Eastern and Southern Africa (COMESA) is a regional trade organization consisting of 20 African states in the eastern and southern regions of Africa and a small number of countries in central and northern Africa. In 2007 its membership included Angola, Burundi, Comoros, Democratic Republic of Congo, Djibouti, Egypt, Eritrea, Ethiopia, Kenya, Libya (since June 2005), Madagascar, Malawi, Mauritius, Rwanda, Sey-chelles, Sudan, Swaziland, Uganda, Zambia, and Zimbabwe. Its objective is to strengthen the institu-tions of member states to help them achieve collective and sustained development.

Origin and Background COMESA, also known simply as the common market, traces its origin to the pan-African vision of economic integration of the African continent that gained prominence between the late 1950s and early 1960s. The consensus then was that the smallness and fragmentation of post-colonial African national markets would constitute a major obstacle to the development of the continent. Accordingly, it was agreed that the newly indepen-dent African states should promote economic co-operation among themselves. In the mid-1960s, the countries of eastern and southern Africa initiated the process toward the formation of an eastern and southern African cooperation arrangement. After the preparatory work had been finalized, the Preferential Trade Area (PTA) Treaty was signed in 1981 and ratified in 1982, giving birth to the Eastern and Southern African PTA.

COMESA was created by treaty in 1994 to re-place the PTA. Its formation was a fulfillment of the requirements of the PTA Treaty, which provided for the transformation of the PTA into a common market. The common market was intended to strengthen the progress of regional integration that had begun under the PTA. The COMESA Treaty had two notable provisions that the PTA Treaty lacked: (1) the concept of multiple speed or variable asymmetry, which allows some countries to progress faster in the regional economic integration process than other countries, and (2) the use of sanctions

(financial penalty, suspension, or expulsion) to discipline member states that fail to implement COMESA programs or to settle disputes arising from the interpretation or implementation of the treaty. These two innovations helped to expedite the process of economic integration in the region.

The affairs of the common market are managed by a number of institutions, including those organs (the Authority of Heads of States and Governments, the Council of Ministers, the Court of Justice, and the Committee of Governors of Central Banks) responsible for making decisions on behalf of the common market; the COMESA Trade and Development Bank, which provides capital for development; the COMESA Re-Insurance Company, which is responsible for providing insurance for trade, investment, and other productive activities, and for promoting trade in insurance and reinsurance business; the COMESA Association of Commercial Banks, which is responsible for promoting and strengthening links between banks in the region; the COMESA Clearing House, which is responsible for settling payments with respect to all transactions in commodities conducted within the common market; and the COMESA Leather Institute, which is responsible for promoting productivity, competitiveness, trade, and regional integration in the leather subsector.

Member countries of the common market are diverse in terms of socioeconomic development and resource endowments. According to 2004 data, gross domestic product (GDP) per capita ranged from as little as US $90 for Burundi to more than US $8,600 for Seychelles. The United Nations Development Program ranked only two countries (Seychelles, followed by Mauritius) as high-human-development countries, six as medium-human-development countries, and 12 as low-human-development countries. The population levels range from less than a million in Comoros to more than 70 million each in Egypt and Ethiopia, with a total of 380 million people in the entire region. Resource endowments vary from agricultural products to crude oil (in Libya) to mineral ores. COMESA was intended to take advantage of the larger market size, to share the region's

common heritage and destiny, and to allow greater social and economic cooperation with the ultimate goal of creating a regional economic community. This regional economic community would then form one of the building blocks on which the creation of the African Economic Community (AEC) would be erected, as envisaged by the Lagos Plan of Action of 1980 and the Abuja Treaty of 1991.

Key Elements and Procedures The COMESA integration strategy comprises the following four phases: (1) the establishment of a free trade area (FTA) through the abolition of all tariff and nontariff barriers on commodities imported from member countries by 2000; (2) the establishment of a customs union with a common external tariff structure by 2004; (3) the adoptions of common investment practices and visa arrangements, and the establishment of a payments union; and (4) the establishment of a common monetary union by 2025. The FTA phase was achieved in 2000 when Djibouti, Egypt, Eritrea, Ethiopia, Kenya, Madagascar, Malawi, Mauritius, Sudan, Zambia, and Zimbabwe eliminated tariffs and nontariff barriers on goods produced in the common market. Burundi and Rwanda eliminated theirs in 2004 and became FTA partners. At the end of 2007, the common market was still operating as an FTA.

The common market has five key provisions. First, there are no custom duties or charges of equivalent effect imposed on goods from one FTA member to another unless it is for protecting an infant industry or against dumping. Second, member states are allowed to impose full national tariff rates on goods from nonmember states. Third, there are no nontariff barriers against goods from one FTA member to another unless such goods are deemed to pose a health or security risk. Fourth, member states follow the COMESA rules of origin in determining whether or not a good is eligible for preferential treatment. Strict local content requirements have been established to prevent nonmember countries from establishing assembly operations in one member country in order to gain duty-free access to the common market. Fifth, member states that are not yet FTA partners but have met the 60 percent tariff

reduction target are granted trade preferences by the FTA partners on the basis of the tariff reductions they have attained. For instance, countries such as Comoros, Eritrea, and Uganda that have reduced their tariff rates on products originating from the common market by at least 80 percent qualify to receive an equivalent reciprocal preferential treatment from the FTA partners. Other COMESA member states that have not implemented the 60 percent minimum tariff reduction do not get any preferential rate from the FTA partners or from those that have reduced their tariffs by at least 60 percent.

Besides market integration, the COMESA integration strategy has been expanded to include transport and communications infrastructure development. More specifically, the focus is on development and implementation of transit traffic facilitation programs; identification and coordination of regional investments in the transportation, communications, and energy sectors; and the promotion and coordination of institutional and policy reforms in the transportation, telecommunication, postal, energy, and environment sectors.

Impact on Member States Economists separate the welfare effects of a regional economic integration into static and dynamic effects. Static effects are short-run effects that relate to productive efficiency and consumer welfare. Static effects are further delineated into trade creation and trade diversion effects, à la Jacob Viner (1950). Trade creation is associated with increase in trade among member states of a regional organization owing to the reduction in tariff and nontariff barriers. Trade diversion occurs when trade with nonmember countries declines as a result of the formation of a regional organization. A regional economic arrangement is welfare-increasing if the trade creation effects more than compensate for the trade diversion effects.

There is evidence of trade creation in the COMESA region. Trade among member countries (intra-COMESA trade) has expanded substantially since the creation of COMESA. Intra-COMESA trade increased from US $1.7 billion in 1994 to US $5.5 billion in 2003. The formation of the FTA has played a major role in facilitating this trade. More

specifically, trade among the FTA partners alone increased from US $1 billion in 2000 to more than US $5 billion in 2005. Jacob W. Musila (2005) investigated the static effects of the COMESA regional trade agreement and concluded that it is welfare-increasing. More specifically, Musila estimated the relative sizes of the trade diversion and trade creation effects for COMESA and found that the trade creation effects exceed the trade diversion effects.

The dynamic effects of regional economic integration relate to the long-run growth rates of the member countries due to increased efficiency as a result of market enlargement. It is believed that large markets permit economies of scale to be realized on certain export goods and, therefore, may lead to specialization in particular types of goods. The net impact of the dynamic effects, like that of the static effects, is not obvious, however. Whether or not regional economic integration increases welfare in the long run depends on the scope of liberalization in member countries and the type of goods in which countries specialize. Athanasios Vamvakidis (1999) shows that growth is faster in economies that liberalize broadly than in those that merely join regional trade agreements and do not liberalize. Augustin K. Fosu (1990) finds that primary commodity exports do not have a significant impact on long-run economic growth.

For COMESA, the approach to regional integration is open regionalism — that is, it liberalizes trade without crowding out the world economy. COMESA member countries specialize mainly in primary commodities, however. As a result, the trade commodity base among member states is narrow and similar. The trade pattern of the common market is such that manufactured goods are imported from outside the region (mainly from rich nations) while primary commodities dominate the exports of the region. And Antonio Spilimbergo (2000) has shown that the importation of manufactured goods may not necessarily result in dynamic gains for less-developed countries (the South). The learning-by-importing models suggest, however, that imports of high-technology goods lead to transfer of technology that

stimulates domestic innovation and economic growth in the importing country.

The experience for the common market region has been a slight improvement in economic growth since the creation of COMESA in 1994. The average growth rate of real GDP per capita per annum increased from about 0.01 percent during 1980 94 to 1.35 percent during 1995 2005. It is likely that the improvement in economic growth in the common market is partly due to the trade liberalization programs implemented under the auspices the CO-MESA Treaty. Together, the static and dynamic effects determine the overall welfare gains or losses associated with regional economic integration. The evidence appears to suggest that there are overall welfare gains in the case of COMESA.

Relation with External Actors COMESA and other African regional trade blocs seek to augment and deepen regional integration on the African continent with a view to creating an AEC. Accordingly, COMESA supports and has signed cooperation agreements with several other African regional undertakings such as the Intergovernmental Authority on Development and the Economic Community of West African States (ECOWAS) and allows overlapping memberships with other regional organizations. Robert Sharer (1999) has observed, however, that overlapping memberships with internal inconsistencies, conflicting regulations and rules, and different strategies and objectives work to impede market expansion and, thus, discourage domestic and foreign investment. Indeed, some of the countries with dual or more memberships, such as Namibia and Tanzania, have been reluctant to implement COMESA programs in full or have quit the common market.

Non-African regional trade blocs are often seen as impeding Africa's regional integration. Jeffrey D. Lewis, Sherman Robinson, and Karen Thierfelder (2003) have observed that North-South trade is more attractive to African countries than South-South trade. Indeed, a majority of the member states of COMESA trade with the European Union (EU) more than with one another. The trade with the EU and the United States was set to increase even further following the opening up of the EU and U.S. markets under the Everything-but-Arms and the African Growth and Opportunity Act initiatives, respectively. A 2001 WTO provision aiming at removing quotas and duties on a large number of goods originating from the world's poorest countries also promised to increase trade with non-African regional blocs and further reduce the advantages that the common market offers to its member states. Ironically, however, the very same processes of global trade liberalization and cooperation with non-African trade blocs have contributed to the evolution of COMESA. The EU directly encourages developing countries to form a group and speak with one voice. For the EU, dealing with a collective organization rather than with numerous individual countries reduces transaction costs. Indirectly, the success of the EU has inspired the hopes for creating the AEC, with COMESA being an intermediate step.

See also customs unions; Economic Community of West African States (ECOWAS); free trade area; regionalism; rules of origin

FURTHER READING

COMESA. COMESA Treaty. Downloadable from http://www.comesa.int/. Offers detailed insights about the es tablishment of COMESA, its objectives, and the various articles and protocols that govern the operation of the organization.

Fosu, Augustin K. 1990. "Export Composition and the Impact of Exports on Economic Growth of Developing Economies." *Economics Letters* 34 (1): 67 71. An em pirical investigation suggesting that exporting of primary products does not have a significant influence on long run economic growth rates of the producing country.

Lewis, Jeffrey D., Sherman Robinson, and Karen Thier felder. 2003. "Free Trade Agreements and the SADC Economies." *Journal of African Economies* 12 (2): 156 206. A general equilibrium analysis of the impact of trade liberalization in the Southern African Develop ment Community (SADC), which shows that trade with the EU is more beneficial to SADC countries than a SADC free trade area would be.

Musila, Jacob W. 2005. "The Intensity of Trade Creation and Trade Diversion in COMESA, ECCAS and

ECOWAS: A Comparative Analysis." *Journal of African Economies* 14 (1): 117 41. An estimation of the static gains/losses of selected examples of regional economic integration in Africa using the gravity model.

Sharer, Robert. 1999. "Trade: An Engine of Growth for Africa." *Finance and Development* 36 (4): 26 29. Ex amines regional integration initiatives and suggests that the overlapping memberships of regional trade blocs in Africa may work to prevent the attainment of dynamic gains.

Spilimbergo, Antonio. 2000. "Growth and Trade: The North Can Lose." *Journal of Economic Growth* 5 (2): 131 46. An analysis of North South trade that suggests that rich nations (the North) may experience tech nological slowdown if they trade with poor nations (the South), and the South could also lose in such trade.

Vamvakidis, Athanasios. 1999. "Regional Trade Agree ments or Broad Liberalization: Which Path Leads to Faster Growth?" *IMF Staff Papers* 46 (1): 42 68. An empirical analysis that suggests that economic growth is faster after liberalizing broadly and slower after joining a regional trade agreement.

Viner, Jacob. 1950. *The Customs Union Issue.* New York: Carnegie Endowment for International Peace. Examines the effects of regional trade agreements and introduces the concepts of trade creation and trade diversion for the first time.

JACOB W. MUSILA

■ comparative advantage

The term *comparative advantage* was first used in England in the early 19th century by economists of the classical school, which dates from the publication of Adam Smith's *An Inquiry into the Nature and Causes of the Wealth of Nations* (1776). Sometimes denoted by the synonymous term *comparative cost*, it expresses the principle by which a nation that opens to international trade is led to specialize in and export certain commodities and to import others. Whereas the former commodities are those in which it has a comparative advantage, the nation is said to have a comparative disadvantage in the latter. The concept of comparative advantage is vital for understanding the structure of world trade and how each country contributes to it. It underlies the whole field of international trade theory and policy, the earliest and arguably the most important of the applied fields of economics and of its predecessor, political economy. Textbooks of international trade devote successive chapters to exploring the reasons why nations have a comparative advantage in certain commodities, and the welfare implications of specialization in accordance with its dictates. Comparative advantage is also relevant to different regions of a country in studying the pattern of interregional trade.

Starting in England at the end of the 17th century with mercantilist writers such as Sir Dudley North and Henry Martyn, the sources of comparative advantage and the gains from the trade that it induces were explored by the luminaries of the nascent economics profession: Smith himself and the classical economists who followed him, such as David Ricardo, Robert Torrens, and John Stuart Mill. Ricardo (1817) enshrined comparative advantage as a key concept for economists by illustrating it with a numerical example relating to trade in wine and cloth between England and Portugal. In the first half of the 20th century the neoclassical school of thought added its own perspectives to the concept of Ricardian comparative advantage when it was generalized by Gottfried Haberler (1936) and critiqued by two Swedish economists, Eli Heckscher (1949) and Bertil Ohlin (1933). Heckscher and Ohlin postulated that the source of comparative advantage resides in the differential factor endowments of trading countries. The Heckscher-Ohlin theory became the mainstream theory of trade after World War II. Since the late 1970s, doubts that had been raised about the empirical validity of comparative advantage led some economists to formulate a New Trade Theory that harks back to another source of trade mentioned by Ohlin (1933), economies of scale, and explains why much of the trade between advanced economies consists of differentiated commodities produced under conditions of imperfect competition.

Despite the frequency with which the principle of comparative advantage is discussed, a precise defi-

nition is hard to find, and most authors cite particular economists or associate the concept with numerical examples, such as Ricardo's, based on two countries, two commodities, and a single input, labor. An algebraic definition along Ricardian lines can clarify the concept. Consider two commodities, 1 and 2, and two countries, A and B. Define L_i^C as the input of labor per unit of output of commodity i ($i = 1, 2$) in country C (C = A, B). Then country A has an *absolute advantage* in commodity 1 if $L_1^A < L_1^B$, and a *comparative advantage* in commodity 1 if

$$L_1^A/L_2^A < L_1^B/L_2^B. \qquad (1)$$

Inequality (1) is consistent with the possibility that $L_1^A < L_1^B$ and $L_2^A < L_2^B$, so that A has an absolute advantage in both commodities but a comparative advantage in the first. Absolute advantage in a commodity thus indicates a smaller absolute cost in terms of a factor such as labor, whereas comparative advantage implies that the ratio between the unit costs of production is lower in A than in B. What is being "compared" in comparative advantage are unit costs relating both to countries and to commodities: a double-barreled comparison. Since relative prices before trade conform to the ratios L_1^C/L_2^C in A and B according to the labor theory of value, the commodity whose relative cost of production is lower is exported in exchange for the other commodity. If $L_1^A < L_1^B$, $L_2^A < L_2^B$, and inequality (1) holds, specialization according to comparative advantage has important welfare implications for both countries: A is better off if it participates in international trade despite its greater efficiency in producing both goods, and B can participate gainfully in the international division of labor despite its absolute disadvantage in both commodities. If $L_1^A < L_1^B$, $L_2^A < L_2^B$, and (1) is changed to the equality $L_1^A/L_2^A = L_1^B/L_2^B$, A has an absolute advantage in both commodities and a comparative advantage in neither. Trade cannot take place since both countries face the same relative price. Comparative advantage clearly trumps absolute advantage as a cause of trade.

According to Jones and Neary (1984, 3), "While the principle of comparative advantage may thus be defended as a basic explanation of trade patterns, it is not a primitive explanation, since it assumes rather than explains inter-country differences in autarkic relative prices." Autarkic relative prices are those that prevail in a state of *autarky*, when a country is closed to trade. Successive models of international trade have been based on different "primitive explanations" for why relative prices differ in autarky, such as intercountry differences in technologies, factor endowments, or tastes. The specification of an autarky equilibrium is, for most countries, a thought experiment designed to ascertain the prices that would prevail in the unlikely event that an economy were isolated or self-sufficient. It is part of an exercise in "comparative statics," in which an economy's resource allocation and economic welfare in autarky are compared to their pattern under free trade. If transportation costs are neglected, free trade equates the prices of traded commodities across countries. Each country's export good rises in value while the import-competing good falls. Under free trade one can therefore no longer speak of a country being a "low-price" country for the goods in which it holds a comparative advantage.

18th-Century Views on the Causes of Specialization Before Ricardo, comparative advantage was interpreted in a looser way than that indicated earlier, namely, as the reason why particular economies show a tendency toward specialization in certain types of commodity. In 1752 the philosopher David Hume published a series of essays analyzing a variety of commercial issues including trade between more developed economies such as England and less developed ones such as his native Scotland. Lower wages give a competitive advantage to more backward economies, allowing them to compete successfully with more advanced ones in the simpler manufactures. In a similar vein the English pamphleteer Josiah Tucker argued that

> It may be laid down as a general Proposition, which very seldom fails, That *operose* or *complicated Manufactures* are cheapest in rich Countries; and *raw Materials* in poor ones: And therefore in Proportion as any Commodity approaches to one, or other of these *Extremes*, in that Proportion it will be found to be cheaper, or dearer in a rich, or a poor

Country. . . . [Moreover] there are certain *local* Advantages resulting either from the Climate, the Soil, the Productions, the Situation, or even the natural Turn and peculiar Genius of one People preferably to those of another, which no Nation can deprive another of. (Tucker 1774, 188, 193)

Tucker not only spoke of national "advantages," but traced them to climate, soil, and even "the natural turn and peculiar genius" of the people, factors that were often subsequently cited as explanations of economic specialization by the classical school of economists. He engaged with Hume in what became known as the "rich country poor country" debate, in which they argued whether poor countries could catch up with or even surpass the standard of living of richer ones. In opposition to Hume, Tucker maintained that the built-in advantages of rich countries are hard to reverse. Without citing Tucker, Adam Smith also anticipated the theory of comparative advantage by maintaining that "the most opulent nations . . . generally excel all their neighbours in agriculture as well as in manufactures; but they are commonly more distinguished by their superiority in the latter than in the former" (Smith 1776, 16). This notion of a "natural" territorial division of labor between rich and poor countries was adopted by most 19th-century economists.

Ricardo's Example of Comparative Advantage

The principle of comparative advantage is contained in chapter 7 of Ricardo's *Principles of Political Economy and Taxation* (1817), which is devoted to foreign trade:

The quantity of wine which she [Portugal] shall give in exchange for the cloth of England, is not determined by the respective quantities of labour devoted to the production of each, as it would be, if both commodities were manufactured in England, or both in Portugal.

England may be so circumstanced, that to produce the cloth may require the labour of 100 men for one year; and if she attempted to make the wine, it might require the labour of 120 men for the same time. England would

therefore find it her interest to import wine, and to purchase it by the exportation of cloth.

To produce the wine in Portugal, might require only the labour of 80 men for one year, and to produce the cloth in the same country, might require the labour of 90 men for the same time. It would therefore be advantageous for her to export wine in exchange for cloth. This exchange might even take place, notwithstanding that the commodity imported by Portugal could be produced there with less labour than in England. Though she could make the cloth with the labour of 90 men, she would import it from a country where it required the labour of 100 men to produce it, because it would be advantageous to her rather to employ her capital in the production of wine, for which she would obtain more cloth from England, than she could produce by diverting a portion of her capital from the cultivation of vines to the manufacture of cloth.

Thus England would give the produce of the labour of 100 men, for the produce of the labour of 80. (Ricardo 1817, 135)

These paragraphs contain four numbers denoting the amounts of labor needed to produce wine and cloth in England (120, 100) and Portugal (80, 90). International trade textbooks interpret them as constant labor coefficients per unit of output of wine and cloth, and deduce from them linear production possibility frontiers and complete specialization in both countries (unless one of them happens to be "large" and remains nonspecialized). Ruffin (2002) and Maneschi (2004) show that this interpretation is incorrect. A close reading of the first three paragraphs of the foregoing passage reveals that the two numbers relating to each country refer instead to the amount of labor embodied in its *total exports* and the amount it would require to produce its *total imports* of the other commodity. In the second paragraph, Ricardo is able to assert which commodity England exports before even mentioning the two numbers relating to Portugal. It is clear that Portugal has an absolute advantage in both commodities since it uses less labor

than England does to produce both traded amounts of wine and cloth. Ricardo did not justify the four numbers by referring to particular economic characteristics of the two countries. Economists have assumed that differences between them in the technology of production account for their respective comparative advantage.

Ricardo never specified the quantities exported by each country. If X is the amount of cloth exported by England and Y that of wine imported, the *terms of trade* (defined as the relative price of the export good) are Y/X units of wine per unit of cloth. Since England requires 100 units of labor to produce X and 120 units to produce Y, the unit labor coefficients are $100/X$ and $120/Y$ respectively, so that its opportunity cost of cloth in terms of wine is $(100/X)/(120/Y) = (5/6)Y/X$. The corresponding unit labor coefficients in Portugal are $80/Y$ for wine and $90/X$ for cloth, so that its opportunity cost of cloth in terms of wine is $(9/8)Y/X$. If these opportunity costs remain constant for all levels of output they are equal to the internal price ratios, so that Portugal has a comparative advantage in wine while England has it in cloth. Since $(9/8)Y/X > Y/X > (5/6)Y/X$, these opportunity costs lie on either side of the terms of trade Y/X, so that trade causes the price of the imported commodity to fall in each country.

Classical Perspectives on Comparative Advantage after Ricardo Another classical economist, Robert Torrens, is often mentioned with Ricardo as a codiscoverer of the comparative advantage principle. In his *Essay on the External Corn Trade* published two years before Ricardo's *Principles*, Torrens argued that even if England produces "corn" (grains) more efficiently than Poland, it may be to England's advantage to import it from Poland: "tracts of her territory, though they should be equal, nay, even though they should be superior, to the lands of Poland, will be neglected; and a part of her supply of corn will be imported from that country" (Torrens 1815, 264 65). Although England has an absolute advantage in corn, it has a comparative advantage in manufactures and specializes in them in order to import corn. Despite Torrens's two-year precedence in print with a passage that shows that he appreciated the dis-

tinction between absolute and comparative advantage, Ricardo deserves his reputation as the discoverer of the principle of comparative advantage since (unlike Torrens) he presented it in terms of a numerical example that allowed both the determination of the direction of trade and the gains from trade accruing to each country. Ruffin (2002) highlights other deficiencies in Torrens's formulation of this principle.

Torrens deserves to be remembered for another reason. In his comparative cost example, Ricardo postulated the existence of a trade equilibrium but never explained how the terms of trade, and hence the division of the gains from trade between the two countries, are determined. Torrens advanced the notion that they depend on the reciprocal demand of each country for the export goods of the other country. As O'Brien (2004) points out, Torrens was a pioneer in the theory of trade policy by proposing that a tariff can turn the terms of trade in a country's favor, and hence insisting on reciprocity with other countries in tariff policy. His argument for an aggressive trade policy was critiqued, and the theory of reciprocal demand developed more rigorously, by John Stuart Mill (1848), who among the classical economists ranks with Ricardo as a great innovator in international trade theory. Mill postulated that reciprocal demand determines the terms of trade in such a way that a country's exports match the value of its imports. He noted that the terms of trade are located between, and bounded by, the autarky price ratios of the two trading countries. The gap between the two countries' autarky price ratios, which indicate their respective comparative advantages, thus fixes the possible range of the terms of trade, while reciprocal demand determines their location in this range. Mill also analyzed how technical change in the export good of one of the countries alters the pattern of comparative advantage and turns the terms of trade against it.

The first professor of political economy in Ireland, Mountifort Longfield, took an important step in generalizing the Ricardian model to many commodities, observing that a commodity is exported by a country if and only if the productivity of the labor

producing it, relative to the other country's, exceeds their relative wages. In his words, "That kind of labour will succeed in each country which is more productive in proportion to its price" (Longfield 1835, 56). Longfield realized the importance of reciprocal demand in determining the range of commodities exported, noting that "if a nation enjoyed an immense superiority in the production of two or three articles of very general demand, the wages of her labourers might be, in consequence, so high that she could not compete with the rest of the world in any other manufacture, under a system of free trade" (69). In the two-commodity Ricardian world, comparative advantage is determined entirely by techniques of production that dictate that one good is exported and the other imported. In a multi-commodity world, a country's comparative advantage depends not only on technology but also on the reciprocal demand of each country for the other's commodities, such that commodities are exported (imported) when the ratios of their labor productivity are greater (smaller) than the ratio of their wage rates. As Longfield asserted, a country that enjoys a very high wage rate compared to its trading partner produces and exports very few commodities.

Neoclassical Perspectives on Comparative Advantage The classical school of thought gradually gave way to the marginalist economics of W. Stanley Jevons in England, Carl Menger in Austria, and Léon Walras in France and Switzerland, and then to the neoclassical school that originated with Alfred Marshall in England. Marshall achieved a "neoclassical" synthesis by combining insights derived from the marginalist economists on the important role played by demand in price determination with the supply-side view of the classical economists that price is determined by the cost of production. This spelled the end of the classical labor theory of value used to determine the autarky prices in the theory of comparative advantage of Ricardo and Mill. The prestige enjoyed by the theory of comparative advantage guaranteed its continued longevity for several more decades, as shown in Jacob Viner's (1937) scholarly *Studies in the Theory of International Trade*, which painstakingly analyzed and critiqued the achievements of the mercantilist, classical, and neoclassical writers in international trade theory and policy.

Marshall did not do much to advance the theory of international trade beyond graphically translating Mill's analysis of reciprocal demand and terms of trade determination by means of the two trading countries' offer curves. The latter show how much of the good in which each country holds a comparative advantage it is willing to export in exchange for alternative amounts of its import good. The slope of the ray joining the origin to the intersection of the offer curves yields the terms of trade. Offer curve analysis was developed in greater detail by Francis Edgeworth, who incorporated the two countries' comparative cost ratios in the offer curves and showed diagrammatically that they bound the terms of trade. Both Marshall and Edgeworth used offer curves to illustrate the impact of a tariff on the terms of trade and to investigate the stability of a trade equilibrium.

Considerably more progress was made by the Austrian-born economist Gottfried Haberler in a 1933 book that appeared in English translation three years later as *The Theory of International Trade with Its Application to Commercial Policy* (Haberler 1936), and by the Swedish economists Eli Heckscher (1949) and Bertil Ohlin (1933). Haberler generalized Ricardian comparative advantage to an economy in which price is defined by the opportunity cost of a commodity in terms of another, rather than by its constant "real cost" of production. He depicted by means of a transformation curve, or production possibilities frontier (PPF), the menu of outputs that can be produced with the economy's factors of production and the available technology. Whereas the PPF is linear in the textbook Ricardian case since real costs of production are constant for any level of output, the neoclassical PPF is concave to the origin. Its slope measures the opportunity cost of one commodity in terms of the other and increases with the level of output because of increasing unit costs. In autarky this slope differs for the two trading partners and signals their comparative advantage in the commodity that is cheaper there. Unlike in the Ri-

cardian case, this slope is no longer independent of demand considerations, and the mix of factors used changes along the PPF together with its slope. Haberler thus generalized the comparative cost model while accepting most of its welfare conclusions.

In an article published in Swedish in 1919 and translated into English in 1949 as "The Effect of Foreign Trade on the Distribution of Income," Eli Heckscher set out to discover for the Ricardian trade model the hitherto missing rationale for why comparative advantage differs across countries and, as the title of his paper suggests, to explore how trade affects income distribution. The clue lies in the difference in the relative abundance of factor endowments across countries combined with differences in the intensities with which commodities use factors in production. Under autarky these features yield differences in commodity prices across countries that in turn lead to trade between them. Making the additional assumption that techniques of production are identical across countries, Heckscher argued that trade causes factor prices to converge and even become equal if neither country becomes fully specialized, a result that later became known as factor price equalization. As in the case of Haberler's PPF, the difference between comparative costs responsible for trade in the first place is erased by the very trade it engenders. After trade is established, and unlike the Ricardian constant-cost case, the difference in comparative costs disappears since it is no longer needed to ensure continued trade.

Haberler's generalization of the Ricardian model and Heckscher's rationale for comparative costs in the context of that model did not go far enough for Heckscher's student Bertil Ohlin, who in 1933 published a book-length critique of the Ricardian model. Together with Heckscher's 1919 article, whose assumptions and conclusions were adopted by Ohlin, it laid the foundations for what became known as the Heckscher-Ohlin theory, the mainstream theory of international trade after World War II. According to Ohlin's critique, the Ricardian model could not be reformed or generalized, but must be rejected as based on false assumptions in order to reconstruct the theory of trade on the new foundations sketched by Heckscher and the neoclassical school. The association of the term *comparative advantage* with the theory of comparative costs that Ohlin wished to discredit explains why this term is hardly used in his 1933 book and disappears completely from the revised edition of 1967. When Paul Samuelson (1948) applied the Heckscher-Ohlin model to two countries, two commodities, and two factors of production and thereby made it more comprehensible and suitable for graphical representation, he was not deterred from using the term *comparative advantage* in connection with it. In fact, Samuelson formulated what became known as the Heckscher-Ohlin theorem: a land-abundant (labor-abundant) country has a comparative advantage in the land-intensive (labor-intensive) commodity and exchanges this commodity for the other since in autarky it is cheaper than in the other country. Its validity depends on numerous assumptions, including identical demands in the two countries such that at the same price ratio they consume goods in the same proportion. It is somewhat ironic that, following Heckscher's lead and despite Ohlin's objection to the term *comparative advantage*, Ohlin ultimately provided the first book-length explanation of the causes of comparative advantage based on international factor endowment differences and differential factor intensities for commodities.

Comparative Advantage and the New Trade Theory The Heckscher-Ohlin (H-O) theorem was tested empirically by Wassily Leontief (1953). Despite his surmise that the United States was then the world's most capital-abundant country, Leontief discovered that it exported labor-intensive and imported capital-intensive commodities! His finding became known as the Leontief paradox and gave rise to numerous articles that attempted to account for it. The underlying H-O theory was extended to more factors than the two (capital and labor) considered by Leontief, including natural resources and skilled labor. Later empirical tests that included these extensions sometimes reversed, sometimes reaffirmed the Leontief paradox. Moreover, contrary to what the H-O theory suggests, most trade flows were shown to occur between industrialized countries whose factor

endowments are fairly similar. Much of this trade is intraindustry in nature, meaning that the same types of commodity (such as automobiles) are both exported and imported. Dissatisfaction with the underlying H-O theory led in the 1970s to the formulation of a *new trade theory*, some of whose models dispense altogether with the notion of comparative advantage. They allow instead for increasing returns to scale, external economies, differentiated products, and the associated imperfectly competitive market structures. Trade can arise even between economies that are identical with respect to factor endowments and technical knowledge. The new trade theorists later realized that Ohlin himself had partly anticipated them in 1933. In chapter 3 of his book, titled "Another Condition of Interregional Trade," Ohlin argued that a powerful secondary reason for trade is economies of scale, due to the indivisibility of certain factors of production. Ohlin remarked that trade and specialization are partly determined by history and accident, factors also stressed by Paul Krugman (1990). Emphasizing increasing returns and monopolistic competition in economies in which comparative advantage plays no role, Krugman showed that the New Trade Theory can provide a satisfactory explanation for intraindustry trade. Other models of the New Trade theorists combine comparative advantage with economies of scale to produce a rich variety of possible trade outcomes.

Comparative Advantage: An Enduring Legacy

Comparative advantage has been attributed to many causes in the trade literature, such as different techniques of production in the Ricardian model and differential factor endowments in the writings of Heckscher and Ohlin. The strategy adopted by the innovators of each theory was to discover a factor that differs between countries, hold everything else in them the same, and build a theory based on this difference. As the French put it, *c'est la différence qui compte*.

Although comparative advantage plays no role in a few models of the New Trade Theory, where the industries that countries adopt are immaterial as long as they end up specializing and thus reaping economies of scale, some of its other models combine such increasing returns with traditional comparative ad-

vantage. Models of the Ricardian or Heckscher-Ohlin type thus endure as significant explanations of comparative advantage in the world economy. The drastic changes in the structure of world trade since the mid-20th century and the emergence of new major exporters such as China and the countries of the "East Asian miracle" present economists and policymakers with the task of analyzing and keeping up with the rapid evolution of comparative advantage in the presence of multinational corporations and of phenomena such as outsourcing and the instantaneous transmission of technological knowledge across frontiers.

See also absolute advantage; economies of scale; gains from trade; Heckscher-Ohlin model; intraindustry trade; monopolistic competition; New Trade Theory; Ricardian model; terms of trade

FURTHER READING

Haberler, Gottfried. 1936. *The Theory of International Trade*. Edinburgh: William Hodge. A comprehensive outline of neoclassical international trade theory and policy by one of its chief contributors.

Heckscher, Eli F. 1949. "The Effect of Foreign Trade on the Distribution of Income." In *Readings in the Theory of International Trade*, edited by H. S. Ellis and L. A. Metzler. Homewood, IL: Irwin, 272 300. First pub lished in Swedish in 1919, the original and remarkably sophisticated presentation of the Heckscher Ohlin the ory of trade.

Jones, Ronald W., and J. Peter Neary. 1984. "The Positive Theory of International Trade." In *Handbook of Inter national Economics*, edited by Ronald W. Jones and Peter B. Kenen. Vol. 1. Amsterdam: North Holland, 1 62. Two noted trade economists survey developments in trade theory from the 1960s to the 1980s.

Krugman, Paul R. 1990. *Rethinking International Trade*. Cambridge, MA: MIT Press. An anthology of articles based on the new trade theory by one of its most eminent exponents.

Leontief, Wassily W. 1953. "Domestic Production and Foreign Trade: The American Capital Position Re ex amined." *Proceedings of the American Philosophical So ciety* 97: 331 49. The article that brought the Leontief paradox to the attention of the profession.

Longfield, Mountifort. 1835. *Three Lectures on Commerce, and One on Absenteeism*. Reprint, New York: A. M. Kelley, 1971. A remarkable anticipation in the early classical period of later developments in international trade theory.

Maneschi, Andrea. 1998. *Comparative Advantage in International Trade: A Historical Perspective*. Cheltenham, UK: Edward Elgar. An analysis of the genesis, evolution, and current significance of the concept of comparative advantage in the theory of international trade.

———. 2004. "The True Meaning of David Ricardo's Four Magic Numbers." *Journal of International Economics* 62 (2): 433 43. A challenge to the conventional interpretation of the numerical example that Ricardo used to illustrate comparative advantage.

Mill, John Stuart. 1848. *Principles of Political Economy*. Reprint, London: Longman, Green, 1920. The most popular 19th century textbook of British classical political economy.

O'Brien, Denis P. 2004. *The Classical Economists Revisited*. Princeton, NJ: Princeton University Press. A thorough analytical survey of the contributions of the classical school of economics.

Ohlin, Bertil. 1933. *Interregional and International Trade*. Revised ed. Cambridge, MA: Harvard University Press, 1967. A comprehensive and literary elaboration of the Heckscher Ohlin theory of trade.

Ricardo, David. 1817. *On the Principles of Political Economy and Taxation*. In *The Works and Correspondence of David Ricardo*, vol. 1, edited by Piero Sraffa. Cambridge: Cambridge University Press, 1951. A fundamental contribution to classical economic theory containing a numerical illustration of comparative advantage and the gains from trade.

Ruffin, Roy J. 2002. "David Ricardo's Discovery of Comparative Advantage." *History of Political Economy* 34 (4): 727 48. An exploration of the path that led Ricardo to the discovery of comparative advantage and a novel interpretation of the numbers he used to illustrate it.

Samuelson, Paul A. 1948. "International Trade and the Equalization of Factor Prices." *Economic Journal* 58 (230): 163 84. An article that expounds and illustrates the Heckscher Ohlin Samuelson theory of trade.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Oxford: Clarendon Press, 1976.

The most famous book of the founder of classical political economy.

Torrens, Robert. 1815. *An Essay on the External Corn Trade*. 1st ed. London: Hatchard. Expounded the notion of comparative advantage in literary form two years before Ricardo's better known numerical example.

Tucker, Josiah. 1774. *Four Tracts, on Political and Commercial Subjects*, tract I. In *Precursors of Adam Smith*, edited by Ronald L. Meek. Totowa, NJ: Rowman and Littlefield, 1973. A preclassical anticipation of the international division of labor and an analysis of the possibility of convergence in the per capita incomes of more and less developed nations.

Viner, Jacob. 1937. *Studies in the Theory of International Trade*. New York: Harper. A magisterial survey of preclassical, classical, and neoclassical trade theories.

ANDREA MANESCHI

■ competition for foreign direct investment
See subsidies and financial incentives to foreign direct investment

■ competition policy
The term *competition policy* is typically taken to include all government policies that influence the degree of competition in a nation's markets, including trade policy. Competition law, in contrast, refers to a specific set of legal provisions concerning the manner in which firms collaborate and compete with one another. That set usually includes measures on cartels, other forms of interfirm collaboration (including so-called vertical restraints), mergers and acquisitions, and abuse of market power. Competition law, therefore, is an element of competition policy but not vice versa.

The appropriate relationship between national trade and competition policies and the merits of different international initiatives on competition law have long been controversial matters both in economic analysis and in international deliberations.

Indeed, if Adam Smith's critique in *The Wealth of Nations* of granting monopolies on trade with Britain's colonies is considered, then these debates are of very long standing. The spread of competition law and the strengthening of existing laws during the current era of international market integration accounts for the renewed interest in the relationship between trade and competition policies.

Many argue that open borders and competition laws are substitutes. Bhagwati (1968) was among the first to show that the pricing power of incumbent firms would be restrained by foreign firms that are prepared to supply domestic customers at world prices. In recent years empirical support for this contention was first provided by Levinsohn (1993) and Harrison (1994). Both showed that the markups of domestic firms declined as tariff rates fell. It should be pointed out, however, that such findings need not imply that setting zero tariffs eliminates market power.

The case for complementing trade reforms with active competition law has been made on several grounds. In the 1980s and 1990s the claim was frequently advanced that prior reductions in tariffs on imports into Japan had not led to greater foreign access to its markets because agreements among Japanese manufacturers, wholesalers, and retailers effectively blocked distribution channels to foreign suppliers (Lawrence 1993; Saxonhouse 1993). More generally, the argument has been made that tariff reductions provide incentives to import-competing firms to take steps, including engaging in anticompetitive practices, that frustrate imports. Private barriers to trade, according to this argument, replace state barriers. A related but distinct argument is that sometimes these anticompetitive practices are facilitated by national unfair trade laws, in particular antidumping measures. The policy implications of the last two arguments are different, however: the former calls for enforcement action against private anticompetitive acts; the latter for reform, if not outright abolition, of unfair trade legislation.

The resurgence of national enforcement actions against international cartels that has taken place since

1993, when the U.S. antitrust authorities offered more generous terms to cartel members to "defect" and turn state's evidence, revealed other complementarities between open borders and competition law (Evenett, Levenstein, and Suslow 2001). Underenforcement of national cartel law creates safe havens where international cartel members can hide evidence, thereby creating a negative cross-border spillover for trading partners. Successful prosecution of an international cartel in one jurisdiction has often prompted investigations in other jurisdictions, creating a positive cross-border spillover. The case for international discipline requiring nations to properly enforce a cartel law can be justified by reference to the first spillover. The case for promoting cross-border cooperation among national competition agencies, which would include the sharing of evidence on cartel investigations, can be made on the basis of the second spillover. Both forms of international collective action would enhance the deterrent effect of national cartel law enforcement.

Since 1995 there has been another wave of cross-border mergers and acquisitions (Evenett 2003). These combinations can involve commercial operations in many jurisdictions and, under current merger review legislation, are often subject to investigations by many national competition agencies. Decisions by national competition agencies to allow or refuse proposed combinations can have cross-border ramifications and from time to time have become the subject of commercial disputes between nations (Muris 2001; Neven and Röller 2000). The adverse reaction of certain U.S. commentators and policymakers to the European Commission's decision on July 3, 2001, to prevent the merger of General Electric and Honeywell is an example. Suboptimal resource allocation is not just a matter of potentially poor analysis of mergers by competition agencies or of different substantive welfare standards, but is also due to the generalized practice of competition agencies of taking into account only those effects of a merger that fall within its jurisdiction's borders. The absence of any compensation mechanism to take account of all relevant cross-

border spillovers, similar to the cross-sectoral bargaining in trade negotiations, is at the center of the inefficiency of simultaneous national merger enforcement. Economic analyses that endogenize the decision to engage in cross-border mergers and acquisitions (Horn and Persson 2001) and that examine the effects of such consolidation in a general equilibrium setting (Neary 2004) have also been developed.

Although considerable attention has been given to competition policy in international forums since the late 1990s, two significant international initiatives predate them. The first was the failed attempt to include binding disciplines on "restrictive business practices" in the postwar multilateral trading system. The chapter on such practices in the Havana Charter, elements of which subsequently provided much of the legal foundation for the General Agreement on Tariffs and Trade, was rejected by the U.S. Congress. Second, in 1980 the members of the United Nations (UN) adopted a nonbinding Set of Multilaterally Agreed Equitable Principles and Rules for the Control of Restrictive Business Practices. The latter remains the only multilateral instrument on competition law.

In July 2004, World Trade Organization (WTO) members decided not to act on proposals to negotiate a multilateral framework on competition policy. The proponents of such a multilateral framework advocated binding obligations to enact and enforce national anticartel law and to adhere to core WTO principles. In addition, provisions to promote both voluntary cooperation between competition agencies and technical assistance for developing countries were advanced (Clarke and Evenett 2003). Some of the strongest opponents to negotiating competition rules at the WTO were the leading competition agencies, fearful of a loss of discretion. These parties created the International Competition Network in 2001, and its numerous activities have contributed to the sharing of best practices in many aspects of competition law. The Organisation for Economic Co-operation and Development has developed a number of recommendations pertaining to competition law and its enforcement. Binding provisions

on competition law have been included in more than 80 free trade agreements, but the evidence to date on these provisions' effectiveness is too sparse to draw any broad policy conclusions.

See also corporate governance; Organisation for Economic Co-operation and Development (OECD); United Nations Conference on Trade and Development; World Trade Organization

FURTHER READING

Bhagwati, Jagdish N. 1968. *The Theory and Practice of Commercial Policy.* International Finance Section, Department of Economics. Princeton, N.J.: Princeton University. Contains an early formulation on the effects that international competition can have on the exercise of domestic market power.

Clarke, Julian L., and Simon J. Evenett. 2003. "A Multilateral Framework for Competition Policy?" In *The Singapore Issues and The World Trading System: The Road to Cancun and Beyond*, edited by State Secretariat of Economic Affairs and Simon J. Evenett. Bern: World Trade Institute. Provides a comprehensive overview of the proposals and criticisms of a multilateral framework for competition policy.

Evenett, Simon J. 2003. "The Cross Border Mergers and Acquisitions Wave of the Late 1990s." In *Challenges to Globalization*, edited by R. E. Baldwin and L. A. Winters. Chicago: University of Chicago Press for the NBER, 411 67. Provides an overview of the huge cross border mergers and acquisitions wave of the late 1990s and the economic questions raised thereby.

Evenett, Simon J., Margaret Levenstein, and Valerie Y. Suslow. 2001. "International Cartel Enforcement: Lessons from the 1990s." *World Economy* 24 (9): 1221 45. Provides both an account of the resurgence of cartel enforcement in the 1990s and an economic analysis of the deterrence effect of national enforcement efforts against international cartels.

Harrison, Ann. 1994. "Productivity, Imperfect Competition, and Trade Reform: Theory and Evidence." *Journal of International Economics* 36 (1 2): 53 73. An early study of the relationship between import competition and market power.

Horn, Henrik, and Lars Persson. 2001. "Endogenous Mergers in Concentrated Markets." *International Jour*

nal of Industrial Organization 19 (8): 1213 44. An analysis that endogenizes the incentive to merge in in ternational markets.

Lawrence, Robert Z. 1993. "Japan's Different Trade Re gime: An Analysis with Particular Reference to *Keiretsu*." *Journal of Economic Perspectives* 7 (3): 3 19. Contends that Japan's industrial structure and potentially anti competitive practices accounts for its low share of im ports of manufactures in GDP.

Levinsohn, James. 1993. "Testing the Imports as Market Discipline Hypothesis." *Journal of International Eco nomics* 35 (1 2): 1 22. Another early and much cited study of the relationship between import competition and market power.

Muris, Timothy J. 2001. "Merger Enforcement in a World of Multiple Arbiters." Prepared remarks before the Brookings Institution's Roundtable on Trade and In vestment Policy (December 21). Washington, DC: Brookings Institution. A thoughtful analysis by a leading antitrust policymaker on the questions that arise when many competition agencies can review a single global merger transaction.

Neary, J. Peter. 2004. "Cross border Mergers as Instru ments of Comparative Advantage." Mimeo. Dublin, Ireland: University College. One of the few analyses of the effects of mergers on the underlying determinants of trade flows.

Neven, Damien, and Lars Henrik Röller. 2000. "The Al location of Jurisdiction in International Antitrust." *Eu ropean Economic Review* 44 (4 6): 845 55. An eco nomic analysis of the effects of many competition agencies reviewing a given merger transaction.

Saxonhouse, Gary. 1993. "What Does Japanese Trade Structure Tell Us about Japanese Trade Policy?" *Journal of Economic Perspectives* 7 (3): 21 43. A critique of Lawrence's viewpoint that recounts the empirical find ings that Japan's trade structure is no different from that predicted by standard theory.

WEB SITES OF POTENTIAL INTEREST:

World Trade Organization. http://www.wto.org/.

International Competition Network. http://www .internationalcompetitionnetwork.org/.

United Nations Conference on Trade and Development. http://www.unctad.org/.

Organisation for Economic Co operation and Develop ment. http://www.oecd.org/.

SIMON J. EVENETT

■ competitive advantage

The concept of competitive advantage came into popular usage in the 1980s as an attempt to identify and define the strategic goals of the firm. In his now famous work *Competitive Advantage: Creating and Sustaining Superior Performance*, Michael Porter showed that the goal of all firms is to achieve a competitive advantage in relation to their rivals. This they do either by selling at a lower cost/price or by differentiating their product from those of their rivals (Porter 1985). Cost and differentiation advantages are frequently referred to as "positional advantages," as they determine the firm's position in an industry, such as whether it competes in the broad, mass market or a segment/niche. A competitive advantage enables a firm to earn profits that are higher than the average profit earned by competitors ("excess prof its"). A *sustainable* competitive advantage is one that cannot be duplicated or imitated by other firms that may enter the industry in the long run and compete for the excess profits enjoyed by existing firms. For example, a firm may acquire a sustainable advantage through the creation of a brand name that comes to be associated with quality or "value for money" and is difficult for other firms to replicate. The St. Michael's label has enabled the British retailer Marks and Spencer to establish a reputation of this kind in the clothing and food retailing industry.

Acquiring a Competitive Advantage Firms ac quire a competitive advantage by creating more value than their competitors in the value chain. The latter is the system of activities (both primary and support activities) that the firm undertakes in the process of value creation. As goods pass through the value chain, value is created at each stage, from upstream suppliers to downstream sales and marketing outlets. Equally important is the contribution made by nonproduc tion activities such as research and development (R&D) and management services. In this process, an

important decision is how to configure each of the activities within the value chain and how to coordinate these activities. Configuration refers to decisions about where to locate each particular activity and how many locations to have for each activity. Coordination relates to the decisions about how to link together the same activity performed in different locations and how to coordinate that activity with other activities in the value chain.

National Competitive Advantage In a later work titled *The Competitive Advantage of Nations*, Porter introduced the notion of national competitive advantage to refer to "the decisive characteristics of a nation that allow its firms to create and sustain competitive advantage in particular fields" (Porter 1990, 18). Such a concept must be distinguished from the concept of comparative advantage used in classical trade theory to refer to the relative cost advantage that a nation enjoys when entering into trade with another country. In classical trade theory, such an advantage was measured in terms of the relative amounts of labor time that producers in different countries required to produce different goods. Later, neoclassical trade theorists explained such differences in terms of factor endowments and the different factor proportions required to produce different goods. While recognizing the usefulness of the concept of comparative advantage for explaining patterns of trade, Porter called for a "new paradigm" based on competitive advantage.

Porter argued that what matters when determining why a particular nation succeeds in a particular industry is the competitive environment of the nation, as this shapes the success of the firms based in that country. This is determined by a number of factors, including the following:

1. *Factor conditions*. These consist of human resources (quantity, quality, and skills of the workers), physical resources (abundance, quality, accessibility, and cost of land and other natural resources, including climatic conditions and location), knowledge resources (scientific, technical, and market knowledge), capital resources (amount and cost of capital to finance industry), and infrastructure (including transportation system, communications

system, health care system, housing, cultural conditions, and quality of life). While neoclassical theory explained national comparative advantage in terms of factor conditions, it assumed that countries were endowed with fixed amounts of these factors. However, although some factors are given (e.g., land and natural resources), countries can add to their stock of other factors through factor creation or investment by individuals or firms.

2. *Demand conditions*. Three broad attributes of domestic demand matter   its composition, its size and rate of growth, and the ways in which it is internationalized. With regard to the composition of domestic demand, the segment structure matters (by giving firms an advantage in segments where domestic demand is strong), as does buyer sophistication (with firms gaining an advantage in sectors where domestic buyers are among the world's most sophisticated and demanding) and the presence of anticipatory buyer needs (domestic buyers are ahead of buyers in other countries). With regard to demand size and the pattern of growth, a large market can enhance competitiveness through economies of scale and learning effects. Also, where demand is rapidly growing, firms may invest more and adopt new technologies more quickly. Finally, it matters how domestic demand is internationalized, pulling a nation's products or services abroad. This can happen as sales of the product to multinational buyers increase and through domestic needs being transformed into foreign needs. The latter can happen through travel and is particularly applicable to cultural goods, such as films or television programs.

3. *Related and supporting industries*. This refers to the presence within a country of internationally competitive supplier industries or other related industries. Porter identified two mechanisms through which a competitive advantage in one industry could benefit other related industries. First, a supplier industry may provide a downstream industry with regular supplies of low-cost inputs and possibly privileged access to these inputs. Second, close working relationships between suppliers and an industry will result in more sharing of information and may result in more rapid innovation within the entire industry.

4. *Firm strategy, structure, and rivalry.* First, the way firms are managed in a particular country and choose to compete matters a great deal. The country's managerial system, the management practices and training, the way in which firms are organized and run, and the extent to which firms are globally oriented and willing to compete internationally are all part of this. Second, the goals of firms within a particular country, along with the motivation of employees and managers, may also contribute to a country's being successful in a particular industry. Third, the more vigorously firms compete with one another in a particular industry, the more likely that a country will develop and successfully sustain a competitive advantage in that industry. Policies designed to create "national champions" by encouraging two or more firms to merge rarely succeed. An important aspect of domestic rivalry is also the ease with which new businesses can be set up in a country and the extent to which they are.

The Porter Diamond Traditionally, these factors are represented by the "diamond" diagram shown in figure 1. In addition to these four factors shaping the competitive environment of a country, Porter emphasized the importance of two other factors that may separately contribute to a country's acquiring a competitive advantage or prevent its doing so. The first is the role played by *chance* events, such as an invention, shortages of key inputs, a sudden rise in the costs of inputs, significant shifts in exchange rates, wars, or the political decisions of government. The second is the role of *government* policy, which may influence firms in a particular industry either positively or negatively. The extent to which government intervention has contributed to the emergence of a national competitive advantage in particular countries is hotly debated. Advocates of government intervention point to the experiences of Japan and Korea as examples of how such intervention can work. By contrast, the same policy largely failed when applied by West European governments to Europe's electronics industry. Government intervention is viewed by Porter as working through the four main determinants of the diamond rather than constituting a fifth factor.
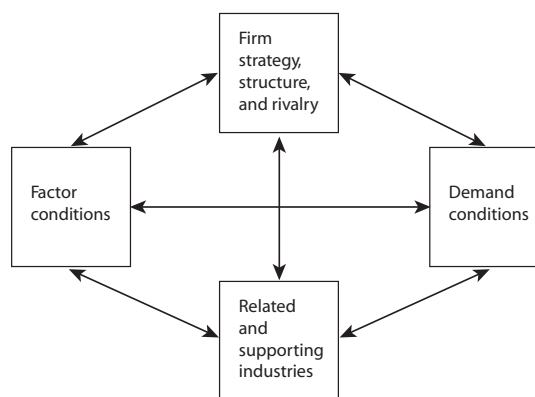
Figure 1
Porter's diamond

An important aspect of the phenomenon of national competitive advantage is that it is an evolving process, in which all the factors interact, with each determinant influencing the others. This process of interaction is as important as the presence or absence of each of the factors viewed separately. Thus factor conditions at any given time are determined by factor creation in the form of investment in education and infrastructure and, more important, investment in the creation and upgrading of advanced and specialized factors such as science and technology. Such investment is stimulated by a strong element of rivalry between domestic firms. Demand conditions are also affected by the existence of intense rivalry, which helps create a large domestic demand for the product through investment in marketing, aggressive pricing policies, and the introduction of more varieties of the same product. The development of related and supporting industries is affected by all of the other factors, as is the degree of domestic rivalry among firms.

Clustering and Sustainability An especially important role in this process is played by the geographical concentration of industries in a particular region, resulting in the phenomenon of clustering. Clustering is a readily observable feature of the economies of most countries. It exists because of the relationships among the four factors in the diamond and the need for firms to be close to exploit these advantages. Within a cluster, the competitive advantage enjoyed

by one industry helps create or sustain the competitive advantage enjoyed by another in a mutually reinforcing process. An example of this is where competitive supplier industries enable a country to sustain a competitive advantage in downstream industries. The same dynamic process may equally well happen in reverse, however.

A further aspect of Porter's diamond model concerns the importance of sustaining a competitive advantage. It is not sufficient for firms in a country to obtain a competitive advantage in a particular industry; that advantage must be sustained by a processing of widening and upgrading. Some of the factors that give rise to a competitive advantage are also important in sustainability. These include investment in institutions that undertake research and generate new ideas, the composition of demand and the intensity of domestic rivalry, and the extent to which the different factors in the diamond interact.

A Multiple Diamond Approach One of the major criticisms leveled at Porter's model is that it fails to adequately address the role of multinational enterprises in the determination of national competitive advantage. In 1993, John Dunning argued that Porter was mistaken in arguing that the firm-specific advantages of multinational companies could be explained purely and simply in terms of the competitive advantages of their home countries (Dunning 1993). Multinational companies derive their advantage from the national diamond existing in a number of countries, not just the home country. If Porter's model is to be applied to international business, Alan Rugman has argued that a "multiple diamond" approach is needed (see Rugman and Verbeke 2001). The weakness of the Porter model, argue Rugman and Verbeke, is that it concentrates entirely on what they call "non-location-bound" firm-specific advantages that are developed by firms in home countries before they engage in foreign direct investment (FDI) and which can then be transferred to other branches of the firm. This ignores, however, the advantages that multinational companies, which have already engaged in FDI, enjoy from having value-added activities in several locations and

from their common governance. Some of these advantages are "location bound," which means that they derive from a particular geographical area and cannot easily be transferred abroad.

The need for such a multiple diamond approach is especially relevant in the case of small countries, where firms are able to overcome the problems of a small domestic market by increasing their sales in the markets of other countries. Rugman and Collinson have illustrated how the Porter model could be applied to the analysis of the problem of international competitiveness with reference to firms in Canada (Rugman and Collinson 2006). Using a double diamond model, Rugman and Collinson show how Canadian firms have used exports to the United States and their rivalry with U.S. firms to develop a global competitive advantage in particular industries. In effect, the Canadian and U.S. markets served as two domestic markets through which Canadian firms could become globally competitive. Such an advantage may be acquired through developing new innovative products and services for consumers in both markets, drawing on support industries and infrastructures in both countries, and making freer and fuller use of the physical and human resources of both countries (Rugman and Collinson 2006).

Competitiveness One of the effects of the focus on competitive advantage since the 1980s has been an increased fascination with competitiveness rankings of countries. Two examples of this are the Global Competitiveness Report by the World Economic Forum (WEF), and the World Competitiveness Center's Yearbook. Such listings seek to rank countries in terms of competitiveness using a range of criteria such as economic performance, government efficiency, business efficiency, and infrastructure. The 2006 rankings of the WEF placed Switzerland, Sweden, and Finland respectively in the top three places in the world. It should be pointed out, however, that it is *comparative*, not *competitive*, advantage that determines what goods a country specializes in. This is because trade is a positive-sum game in which all countries gain, not a zero-sum game in which one country gains only at the expense of another.

This point is well made by Paul Krugman in a short essay first written in 1994, in which he comments on what he called the "dangerous obsession" with competitiveness (Krugman 1996). Using the example of Japan achieving a higher level of productivity than the United States, Krugman writes, "While competitive problems could arise in principle, as a practical, empirical matter, the major nations of the world are not to any significant degree in economic competition with each other. Of course, there is always a rivalry for status and power   countries that grow faster will see their political rank rise. So it is always interesting to *compare* countries. But asserting that Japanese growth diminishes US status is very different from stating that it reduces the US standard of living   and it is the latter that the rhetoric of competitiveness asserts" (Krugman 1996, 10).

Achieving a national competitive advantage is important in enabling some countries to grow faster than others. An obsession with competitiveness, however, is dangerous, as Krugman argues, because it can lead to protectionism and trade conflict, as well as misguided public spending designed to create and sustain a national competitive advantage over other trading partners and the promotion of national champions.

To summarize, acquiring and sustaining a competitive advantage can be seen as the primary strategic aim of firms in the modern world economy. The source of national competitive advantage is to be found in the competitive advantage obtained by the firms operating in the country, both domestically owned and foreign owned. This, in turn, is the product of a complex set of factors that interact with one another to shape the competitive environment of the country. Unlike comparative advantage, however, competitive advantage is a dynamic concept, in which national competitive advantage continuously evolves. In a world in which multinationals are dominant, the source of such advantage often results from firms operating in more than one country. Despite its usefulness in explaining why some countries develop a specialization in a particular industry or branch, an emphasis on attaining a competitive advantage can cause governments to pursue

misguided policies that lead to trade conflict. It remains the case that trade is a positive-sum game in which all can benefit.

See also commodity chains; comparative advantage; foreign direct investment (FDI); foreign direct investment: the OLI framework; New Economic Geography

FURTHER READING

Dunning, John H. 1993. "Dunning on Porter." In *The Globalisation of Business*, edited by John H. Dunning. London: Routledge. A useful and interesting paper criticizing Porter for not specifically including multina tional activity as an element in the model shaping the competitive environment of a country and arguing for an extension to take into account multinational enterprises whose competitive advantage is derived from operating in several countries.

Krugman, Paul. 1996. *Pop Internationalism*. Cambridge, MA: MIT Press. A collection of short essays on the dangers of some popular ideas about trade, with the first chapter being a critique of the contemporary "obsession" with the notion of competitiveness and subsequent chapters further developing this theme.

Porter, Michael E. 1985. *Competitive Advantage: Creating and Sustaining Superior Performance*. New York: Free Press. An earlier work by Porter setting out and ex plaining the notion of competitive advantage as applied to firm strategy. It introduces the idea of generic strategy and distinguishes between competitive scope and com petitive advantage.

   . 1990. *The Competitive Advantage of Nations*. London: Macmillan. A book presenting the basic "the ory" of competitive advantage developed by the author and containing the results of extensive studies of com petitive advantage in relation to both individual indus tries and countries. The final section examines the im plications of competitive advantage for company strategy and government policy.

Rugman, Alan M., and Simon Collinson. 2006. *Interna tional Business*. 4th ed. Harlow, UK: Pearson Education. A leading textbook on international business containing a useful chapter titled "Corporate Strategy and National Competitiveness," which explains and evaluates the Porter diamond model and calls for a multiple country diamond model. A particular attraction of the chapter is

the use of several case examples to illustrate how this operates in practice.

Rugman, Alan M., and Alain Verbeke. 2001. "Location, Competitiveness, and the Multinational Enterprise." In *The Oxford Handbook of International Business*, edited by Alan M. Rugman and Thomas L. Brewer. Oxford: Oxford University Press, chapter 6. A chapter concerned with the influence of locational factors on international competitiveness, including an assessment of Porter's contribution to this through his work on national competitive advantage and introducing the authors' own conceptual framework. The latter dispenses with the single home base perspective of Porter's analysis and emphasizes the importance of foreign locations for the overall competitiveness of multinationals.

NIGEL S. GRIMWADE

■ competitive devaluation

See beggar-thy-neighbor policies

■ compulsory licensing

See access to medicines

■ computable general equilibrium models

See applied general equilibrium models

■ conflicted virtue

In the world's monetary system, only a few major currencies mainly the U.S. dollar and the euro dominate as units of account for the internationally liquid assets or debts of developing countries or emerging markets. The consequences of this international asymmetry among currencies are profound. It leads to the problems of original sin (the inability of a debtor country to borrow in its own currency) and conflicted virtue (the inability of a creditor country to lend in its own currency).

Any creditor country that cannot lend in its own currency cumulates a currency mismatch called con-flicted virtue a concept introduced by McKinnon (2005). A country is "virtuous" if it has a high saving rate and tends to run surpluses in the current account of its balance of payments, that is, make loans to foreigners. Because the surplus country cannot lend in its own currency, however, it builds up claims on foreigners in foreign exchange, usually U.S. dollars. This situation has led foreign debtor governments particularly the U.S. government to complain that the surplus country's ongoing flow of trade surpluses and official reserve accumulation of dollars is an unfair result of having an undervalued currency. Debtor governments may then try to pressure the surplus country to appreciate its currency against the dollar in the often mistaken belief that an appreciation of the surplus country's currency will reduce its trade surplus.

The greater the foreign mercantile pressure for appreciation, the greater the concern of domestic private holders of dollar assets that they will suffer capital losses in their home currency. As holders of dollar assets switch into the domestic currency and out of dollars, the government is "conflicted." An appreciation would dampen exports and, if expected to be repeated, induce serious deflation with a zero percent interest liquidity trap (a situation in which the nominal interest rate is close to zero and the country's central bank is unable to stimulate the economy with monetary policy), as in Japan in the 1990s. But the American government may threaten trade sanctions if the creditor country under siege does not allow its currency to appreciate substantially. This is the essence of the syndrome of conflicted virtue.

Conflicted Virtue in East Asia Conflicted virtue would not arise in creditor countries whose money is the dominant vehicle currency in international finance. In the 19th century, Britain was the largest creditor country, with large current account surpluses financed by large net capital outflows. Most of the British claims on foreigners, however, were denominated in the pound sterling, the British currency. Thus British investors were happy to accumulate large sterling claims on foreigners without provoking a depreciation of their foreign claims

against their domestic ones. Similarly, for 25 years after World War II, the United States had trade surpluses and was the world's biggest creditor. By then, however, the dollar had displaced sterling as the world's dominant vehicle currency. Because U.S. claims on foreigners were denominated mainly in dollars, there was no internal currency mismatch in U.S. private portfolios.

The East Asian economies (and, increasingly, oil-rich Middle Eastern countries) are unusual, however, in that they are significant international creditors whose currencies are relatively little used outside their own countries. In Japan, large current account surpluses have persisted since the late 1970s. Taiwan's and Singapore's current account surpluses have been significant since the 1980s. Since 1998, previous debtor economies such as Korea have run current account surpluses (reflecting their "virtuously" high saving rates), resulting in high net capital exports. Sometimes these are offset by foreign direct investments (FDI) abroad. But the common mode of finance is to build up liquid claims either privately or as official exchange reserves in international monies such as the dollar. Japan's current account surpluses continued into the new millennium. By 2005, Japan's net holdings of liquid claims on foreigners (largely in dollars) reached a new high of U.S. $1.7 trillion, of which about $830 billion were official exchange reserves (McKinnon and Schnabl 2006).

Although China's buildup of liquid dollar claims has a much shorter history than Japan's, it has been accentuated by large inflows of FDI relatively illiquid long-term liabilities. By the end of 2005, China's liquid dollar-denominated assets were roughly $1 trillion, of which $819 billion were official exchange reserves a higher proportion than in Japan. The cumulative joint holding of dollar claims of all East Asian countries amounts to nearly $3 trillion. In China, household consumption, wages, and claims on financial intermediaries such as banks (deposits) and insurance companies (annuities) are mainly in yuan. Thus private Chinese households and firms will hold dollar assets only if there is a substantial business convenience in doing so, or if the interest rate on dollar assets is higher, and if the im-

mediate threat of yuan appreciation (dollar devaluation) is absent.

Effects of Conflicted Virtue China and other East Asian governments worry about the sudden loss of export competitiveness should their currencies be forced to appreciate. Beyond this, when the world price level measured in dollars is stable, any such appreciation would be followed by a domestic deflationary spiral as in Japan from the mid-1980s through the 1990s as a result of the erratically appreciating yen. McKinnon and Ohno (1997) describe the nature of an American mercantile pressure to appreciate the yen from the 1970s to the mid-1990s, then analyze the subsequent deflationary consequences for Japan.

Governments in creditor economies with conflicted virtue may cut domestic short-term interest rates to forestall or slow the conversion of privately held dollar assets into domestic currency. Insofar as people believe that low short-term rates will persist, domestic long-term interest rates also fall. At any given exchange rate, a new portfolio equilibrium can be found in which private agents would be willing to finance the nation's ongoing current account surplus by building up higher yield liquid dollar claims rather than the government accumulating official exchange reserves.

Japan has the longest experience with current account surpluses, and the associated buildup of dollar claims, from the early 1980s into the new millennium. Between 1978 and 2007, the interest rates on long-term (10-year) Japanese government bonds (JGBs) averaged about 3 to 4 percentage points less than those on long-term U.S. treasuries. Short rates are more volatile, but were bid down close to zero in the mid-1990s. In November 2006, when the American interest rate on overnight federal funds was 5.25 percent, the corresponding Gensaki rate in Tokyo was just 0.25 percent. As long as interest rates on yen assets are sufficiently below those on dollar assets, the Japanese private sector banks, insurance companies, trust funds, and so on can still be persuaded to fund Japan's ongoing current account surpluses by accumulating dollar assets with higher yields. This interest differential reflects both the ex-

pectation that the yen will appreciate on an upward trend, as was important from the late 1970s to the mid 1990s, and a negative risk premium that arises when the yen simply fluctuates against the dollar so as to make the ever-larger private holdings of dollar assets more risky in yen terms as was the case in the first few years of the new millennium.

But there are limits on how negative this risk premium on yen assets, and on how wide the associated interest differential, can become. When American interest rates fell to abnormally low levels, with short-term rates down to just 1 percent in early 2004, the spread was not (could not be) big enough because Japanese nominal interest rates were bounded from below by zero. Then the Japanese private sector refused to keep acquiring enough dollar assets to finance the current account surplus. Indeed, private agents in Japan started dishoarding previously accumulated dollar assets in order to acquire near-zero-yield yen assets! In 2003 4, the Bank of Japan intervened massively in the foreign exchange markets to acquire more than $330 billion mostly from private Japanese financial firms in order to prevent the yen from appreciating sharply.

Since 2004, China has run a larger current account surplus, mainly financed by a huge buildup of official exchange reserves. This buildup has led to calls from the American government for China to appreciate its currency the threat thereof leading to a fall in internal Chinese interest rates and the unwillingness of private Chinese wealth holders to acquire overseas assets. Therefore, the rapid accumulation of official exchange reserves becomes even larger. Conflicted virtue in creditor countries, with their overhangs of dollar assets, continues.

See also balance of payments; Bretton Woods system; carry trade; dollar standard; dominant currency; foreign exchange intervention; global imbalances; gold standard, international; interest parity conditions; International Monetary Fund (IMF); international reserves; mercantilism; original sin; reserve currency; vehicle currency

FURTHER READING

McKinnon, Ronald. 2005. Exchange Rates under the East Asian Dollar Standard: Living with Conflicted Virtue. Cambridge, MA: MIT Press. Focuses on East Asian exchange rate and monetary policy, with specific reference to Japan, China, and regional monetary cooperation. It elaborates on the concept of "conflicted virtue" as applied to East Asia.

McKinnon, Ronald, and Kenichi Ohno. 1997. Dollar and Yen: Resolving Economic Conflict between the United States and Japan. Cambridge, MA: MIT Press. Shows how forced appreciations of the yen in the 1980s to the mid 1990s deflated the Japanese economy and undermined subsequent economic growth.

McKinnon, Ronald, and Gunther Schnabl. 2006. "China's Exchange Rate and International Adjustment in Wages, Prices, and Interest Rates: Japan Déjà Vu?" CESifo Studies 52 (2) (June): 276 304. Analyzes the impact of China's exchange rate changes on the domestic economy and draws parallels with the Japanese experience.

RONALD I. McKINNON

contagion

In economics, *contagion* refers to the spread of a crisis from one country to others. Such a spread can show up in a variety of ways, such as its effects on interest rates, asset prices, exchange rates, capital flows, or the probability of a crisis. Economic usage of the word *contagion* draws on the analogy to the spread of contagious diseases. Thus the currency crisis in Thailand that spread to much of the rest of East Asia in 1997 was often referred to as the "Asian flu" or the "Asian contagion." The use of the term *contagion* became particularly popular with the crisis in Mexico at the end of 1994, which was felt throughout most of Latin America and became known as the "tequila crisis." Academic attention to contagion had been stimulated by the earlier crises in the European Monetary System in 1992 and 1993.

Defining Contagion There is considerable debate among economists about how to define contagion. Some hold that it refers to any general transmission of shocks from one market to another. Others argue that contagion exists only if there is a significant increase in cross-market relationships beyond what can be explained by the fundamental state of a country's

economy ("fundamentals"). Still others contend that contagion spreads only through certain channels. In an influential analysis, the economist Paul Masson (1999) defined contagion as a consequence of sudden shifts in market expectations and investors' confidence. Such shifts are particularly relevant in the context of second-generation crisis models in which a country's fundamentals can be in an intermediate vulnerable zone between good and bad. In such situations, a crisis in one country can directly affect expectations of the situation in other countries, and hence move an economy from a good equilibrium to a bad one and end in a crisis. Masson distinguishes contagion from other types of shocks. He calls a common external shock that affects a group of countries "monsoonal effects," referring to the example of a monsoon season affecting several countries in a region. This type of effect, Masson argues, should not be considered contagion. Another example of a monsoonal effect is a shock to oil prices or supply that affects many countries at the same time. Still another is the effect on developing countries of major changes in the interest rates of large industrial countries. Both of the latter two examples helped contribute to the Latin American debt crises of the 1980s.

During noncrisis periods, contagion is usually related to the linkages of the economic fundamentals between one or more countries, and is typically called *interdependence*. Possible factors leading to contagion include trade linkages, common external shocks, and political or economic policies. Contagion is more frequent during crisis periods than noncrisis periods, however, and is an important explanation of emerging market crises of the late 20th and early 21st centuries. Crisis contagion beyond what can be explained by fundamental links is often called "pure contagion." "Unjustified" contagion refers to pure contagion resulting from panic and other forms of irrationalities and market imperfections. No commonly accepted name has yet been given to rational pure contagion that is transmitted through financial channels such as portfolio rebalancing. A country can be an "innocent victim" of fundamental-based contagion or of pure contagion. For example, the con-

tagion effect from Argentina to its small neighbor Uruguay in 2001 was due primarily to economic interdependence, while the spread of the Asian crisis has often been attributed to panic contagion. The degree of a country's innocence and the degree to which contagion is unjustified by fundamentals are often matters of considerable dispute.

Causes and Channels of Contagion Contagion can be transmitted through real or financial linkages, or through other channels such as global investors' behavior and market imperfection. Causes of contagion are often closely related to its transmission channels. Real linkages, such as trade, economic policies, shared fundamentals, and common shocks, all lead to increased interdependence among economic and financial markets. Contagion effects generally have a strong regional concentration (the Russian crisis in 1998 is a major exception). This has led many economists to stress the role of direct macroeconomic linkages, especially through international trade, as an important channel for contagion.

If contagion is taken to include the general interdependence among economic fundamentals, it can be measured using comovements of financial or economic variables. Pure contagion measurements typically test for cross-country relationships beyond fundamentals. Treatments of what are considered fundamentals also vary substantially. Typically, the broader the range of factors that analysts consider fundamentals, the less pure contagion they see.

Contagion also can spread through financial connections. Global portfolio diversification allows large capital flows to enter emerging markets but also makes inflows quickly become outflows during a crisis. Such abrupt reversals of capital flows are often called "sudden stops." In order to maintain the liquidity of their portfolios, institutional investors tend to sell their cross-border investments in one market when another market is in trouble. Portfolio diversification and rebalancing alone are often enough to explain contagion effects without invoking market imperfections. Similarly, common lender problems also lead banks to rebalance their assets portfolio and refuse to roll over debt to countries with risk characteristics similar to those countries in

default. If several countries share the same type of investors, correlations among these countries' financial assets increase, even among countries with few economic similarities. If investors trade only certain categories of assets, the probability of selling out that type of asset during a crisis also increases.

Financial market imperfections, such as asymmetric information, a situation in which one party has more information than another, and ill-designed incentive structures for fund managers, can also contribute to the spread of a crisis from one market to another. When a market is less transparent and information is not readily available, investors are more likely to herd. When investors are relatively uninformed they may assume that if someone else takes an action, it is because that person has better information. Such behavior would increase the size of market swings.

Fund managers' compensation contracts, even if optimal at the firm level, may lead to inefficiencies at the macroeconomic level. Fund managers, for example, are usually evaluated on short time frames, which restrict them from taking advantage of mispricing. Moreover fund managers are typically evaluated against a benchmark, encouraging them to form a portfolio very close to that benchmark or to follow the actions of other fund managers. Such "separation of brain and resources" contributes to fund managers' pursuing short-term profits and herding, which can cause cross-border contagion.

Owing to similarities among the countries, their geographic proximity, and limited information, investors may not always be able to differentiate among common shocks and country-specific shocks. As a result, they might make heuristic decisions. Herding and rational ignorance are such "rule of thumb" responses to market imperfections, and contagion is the consequence.

Investors' sentiments   panic, overoptimism, overreaction, and shifts in expectation   can also lead to contagion. During a crisis, investors might panic and pull out funds in the entire crisis region regardless of each country's fundamentals. They can be overly optimistic before a crisis. A crisis in one country might force them to reevaluate investments that they

initially should have investigated more thoroughly. Thus a crisis in one country can serve as a wake-up call prompting a reevaluation of investments in other countries with shared characteristics.

Of course, if markets were already fully efficient there would be no need for such reevaluations. However, many economists and finance experts are now giving much greater attention to ways in which financial markets may be less than ideally efficient. Such analysis is called behavioral finance. It varies from emphasis on external constraints such as the high cost of information to possible psychological biases. Behavioral analysis is still at an early stage of development, but it promises a number of fruitful insights into the behavior of actual financial markets that fall between the extremes of ideally efficient markets and wildly irrational ones (see Willett 2000).

Examples of Crisis Contagion Discussions of contagion often suggest that the spread of crises is especially associated with the effects of panic in financial markets. This panic contagion view has become particularly widespread in light of the 1997 98 Asian crisis as the countries hit had strong macroeconomic fundamentals. Subsequent analyses, however, showed that many of the countries in crisis had microeconomic and financial weaknesses that suggested some justification for the spread.

Since many crisis countries in Asia had impressive economic growth records, balanced government budgets, and conservative monetary policies, some economists started to analyze the role market psychology and investors' behavior played during the crisis. South Korea is a prime example. Despite its good macroeconomic fundamentals, Korea suffered devastating losses during the crisis largely because of financial and microeconomic weaknesses. Because of implicit guarantees against large depreciation of the exchange rate and the prospect of government bail out of failing institutions, the banking sector in South Korea had incentives to engage in excessively risky loans. As a result it became burdened with nonperforming loans while the corporate sector took on large dollar-denominated, short-term debt. Similarly, a large number of Indonesian and Thai corporations had been borrowing heavily in U.S.

dollars. They and their overseas investors had believed that their countries' exchange rates were safe from large depreciations. When this assumption proved false, there was a rush to cover their positions, which in turn generated large capital outflows.

Furthermore, while many initial analyses treated the whole set of crises as a single event, the crisis developed over a period of months and had several distinct phases. The first phase, between July and October 1997, started with the run on the Thai baht. Its subsequent depreciation spread mainly to Indonesia, Malaysia, and the Philippines, all of which were forced to allow large depreciations of their currencies. The financial markets of a much broader group of countries felt ripples, but did not experience strong speculative attacks or capital flight. In the second phase, which started in October 1998, the devaluation of the Taiwan dollar generated fears that the Hong Kong dollar would follow. The Hong Kong stock market lost 40 percent of its value, which led to the devastating devaluation of the South Korea won in less than two months. While Korea felt some immediate repercussions from the Thai crisis, they were relatively mild. The strong speculative attacks on the won did not occur until months later. Simple panic contagion would not explain such a protracted spread.

The Asian financial crisis drove down raw material prices. This had severe consequences for Russia, which was highly dependent on exports of raw materials. Even though interest rates soared to 150 percent to attract investors to buy government bonds, by mid-1998 Russia was in need of help from the International Monetary Fund (IMF) to maintain its exchange rate. However, the Russian government failed to implement a realistic budget and necessary legislation to meet IMF requirements. Fearing the IMF might pull the plug on its loans, global investors continued to flee Russia, and Russia soon after defaulted on its debt, sending shock waves through financial markets across the globe.

The financial market expected a substantial devaluation of the ruble, but most observers believed that Russia was too important geopolitically to be allowed to default. Thus markets were stunned.

Combined with the near collapse of the hedge fund Long-Term Capital Management in the United States, the Russian default set off a shift to extreme risk aversion among investors. Risk premiums rose substantially, and emerging market countries found it difficult or impossible to find investors for new international credit instruments for many months.

Policy Implications The best ways to prevent contagion and to deal with it when it occurs vary depending on its causes and channels. A number of policies that are good in their own right, such as the promotion of sound economic fundamentals and the development of better policies in both the public and private sector for assessing and managing risk, also help reduce problems from inflation. The role played by contagion during the crises of the 1990s prompted many proposals for greater national and international control over international financial flows.

Confusion has resulted from the tendency of the popular press to use the term *contagion* to describe any effects of a crisis in one country on currency or financial markets of other countries. The crisis in Iceland in 2006 was felt in currency markets as far away as Eastern Europe and Africa, but these fairly mild ripple effects from Iceland were of a quite different magnitude than the devastating crises that hit a number of Asian countries after Thailand's currency crisis. The global repercussions of the Russian default in 1998 fell between these extremes.

The available evidence clearly suggests that major shocks often do cause indiscriminant contagion in financial markets, but with the exception of the Russian default this tends to last for only a brief period, measured in hours, days, or weeks. Medium-term responses from the currency and financial markets tend to be much more differentiated, although the determinants of these more focused medium-term responses can be quite complex, including both trade and financial linkages and a broad range of fundamentals including financial and political as well as economic considerations. Since information is often quite imperfect, perception of the fundamentals can shift without any change in the actual fundamentals themselves. Thus a crisis in one

country can act as a wake-up call that generates re-evaluations of conditions in other countries. The Thai crisis provides a vivid example.

In general, the fallout on other countries and markets from the crises occurring since the Russian default have been much milder, and as a consequence support for major reforms in the international financial architecture has fallen off. Many economists, however, still believe that there is a strong case for strengthening the capabilities of the IMF to act as an international lender of last resort to deal with liquidity crises, especially those that would have the potential to generate considerable contagion.

See also asymmetric information; banking crisis; capital flight; capital flows to developing countries; currency crisis; financial crisis; hedge funds; hot money and sudden stops; international financial architecture; International Monetary Fund (IMF); International Monetary Fund conditionality; International Monetary Fund surveillance; Latin American debt crisis; lender of last resort; sovereign risk; spillovers

FURTHER READING

Claessens, Stijn, and Kristin Forbes, eds. 2001. *International Financial Contagion: How It Spreads and How It Can Be Stopped.* Boston: Kluwer Academic Press. Examines the causes, channels, and measurement issues of international financial contagion.

Eichengreen, Barry. 1999. *Toward a New International Financial Architecture: A Practical Post Asia Agenda.* Washington, DC: Institute for International Economics. Critically reviews proposals for reform of the international financial system and national policies to reduce the incidence of currency crises and contagion.

Masson, Paul. 1999. "Contagion: Monsoonal Effects, Spillovers, and Jumps between Multiple Equilibria." IMF working paper 98/142. Washington, DC: International Monetary Fund. Usefully distinguishes several concepts of contagion.

Rajan, Ramkishen S. 2003. "Unilateral, Regional, and Multilateral Options for East Asia." In *Financial Governance in East Asia*, edited by Gordon De Brouwer and Yunjong Wang. London: Routledge, 239 63. Considers a list of policy options to deal with financial crises and contagion.

Willett, Thomas D. 2000. "International Financial Markets as Sources of Crisis or Discipline: The Too Much, Too Late Hypothesis." Princeton Essays in International Finance No. 218 (May). Princeton, NJ: Princeton University. Discusses the usefulness of behavioral finance for analyzing the behavior of international financial markets that are neither fully efficient nor wildly irrational.

Willett, Thomas, Ekniti Nitithanprapas, Isriya Nitithanprapas, and Sunil Rongala. 2005. "The Asian Crises Reexamined." *Asian Economic Papers* 3 (3): 32 87. Analyzes hypotheses and evidence for the causes of the Asian crisis and reviews the contrasting views on the role of contagion in the Asian crisis. Offers extensive references.

PRISCILLA LIANG AND THOMAS WILLETT

■ Convention on Biological Diversity

The Convention on Biological Diversity (CBD) aims to promote the well-being of both nature and humans. Negotiated under the auspices of the UN Environment Program (UNEP), adopted on May 22, 1992, and entered into force on December 29, 1993, it is built on a threefold, interacting objective: "the conservation of biological diversity, the sustainable use of its components and the fair and equitable sharing of the benefits arising out of the utilization of genetic resources" (art. 1). The convention is comprehensive in scope, as it includes all species and ecosystems worldwide, as well as the genetic diversity within species. There are currently 188 parties to the convention, which is governed by a conference of the parties (COP) that is supported by a Subsidiary Body on Scientific, Technical, and Technological Advice. The CBD secretariat is located in Montreal and has about 60 professionals and support staff.

The background for negotiating and strengthening the CBD was an increasing awareness and agreement among scientists that the current rate of species extinction was (and remains) extremely high compared to the natural average rate (Wilson 1988; Heywood 1995, 232). A rapid loss of genetic diversity in domesticated plants, with risks of reduced food security, was another central concern. Of the

estimated 7 to 30 million species on earth, only 1.9 million have been scientifically described   most of what is lost is scientifically unknown. Moreover, developments in the biotechnologies have increased economic interests in the world's genetic resources. Genetic resources constitute important input factors in biotechnology, including the pharmaceutical sector and all development and breeding of plants and animals. Wild relatives of domestic crops provide genetic variability that can be crucial in circumstances such as overcoming disease outbreaks or adjusting to climatic changes. Species may contain compounds (genes, proteins) that can generate valuable pharmaceuticals or other products at some future date (Laird and Wynberg 2005).

Provisions on access to genetic resources, including the equitable sharing of the benefits of their use, form a central element of the CBD and are regarded as a prerequisite for the two first objectives of conservation and sustainable use. The CBD reconfirmed national sovereign rights to genetic resources (art. 15.1) and equitable sharing of benefits from use of those resources (art. 15.7). Access to the resources is based on mutually agreed terms and subject to prior informed consent (art. 15.4 and 15.5).

The value of products derived from genetic resources worldwide has been estimated at between U.S. $500 billion and $800 billion (Ten Kate and Laird 1999). In addition to the direct economic values of genetic resources, biodiversity as a whole provides a great range of ecosystem services, such as local water and climate regulation, and materials for building and firewood. There is also a great range of noneconomic values attached to biodiversity, such as cultural and intrinsic values. The average annual loss of wild habitats and populations is estimated to deprive people of ecosystem services with a net worth of about U.S. $250 billion every year (Millennium Ecosystem Assessment 2005). The CBD parties have pledged to halt and reverse the loss of biodiversity by 2010.

To this end, the CBD parties have agreed to develop national biodiversity strategies, integrate biodiversity conservation in all policy levels and sectors, identify and monitor biodiversity, establish systems of protected areas, and identify activities that are likely to have adverse effects on biodiversity. Moreover, the parties must adopt economically and socially sound measures to act as incentives for conservation and sustainable use, establish programs for scientific and technical education and training for identification and conservation, and provide support for such training in developing countries. The CBD is equipped with a monitoring mechanism in the form of national reporting and an incentive mechanism in the form of the Global Environment Facility (GEF). The GEF has invested nearly U.S. $4.2 billion for biodiversity conservation in developing countries since its establishment in 1991.

In several respects, the CBD constitutes a framework agreement open for further developments and specifications. For example, article 19.3 commits parties to the CBD to consider the elaboration of a protocol to protect biodiversity from potential risks from genetically modified organisms. On this basis, the parties negotiated the Cartagena Protocol on Biosafety, which was adopted in 2000 and entered into force in 2003. The protocol builds on a precautionary approach and establishes an advance informed agreement (AIA) procedure to help countries make decisions before agreeing to the import of such organisms into their territory.

The COP has initiated work on seven thematic work programs. These address marine and coastal biodiversity, agricultural biodiversity, forest biodiversity, island biodiversity, the biodiversity of inland waters, dry and subhumid lands, and mountain biodiversity. In addition, work has been initiated on biosafety; access to genetic resources; traditional knowledge, innovations, and practices (art. 8[j]); intellectual property rights; indicators; taxonomy; public education and awareness; incentives; and alien species.

The ecosystem approach is the primary framework for action under the convention. It is a strategy for integrated management of land, water, and living resources that aims to promote conservation and sustainable use in an equitable manner. The ecosystem approach recognizes that humans, with their cultural diversity, are an integral component of eco-

systems. As the greatest threat to biodiversity lies in the replacement of undeveloped land by alternative systems of land use, the approach recommends management in an economic context in order to reduce market distortions, which undervalue natural systems and provide perverse incentives and subsidies.

The COP has requested close collaboration with relevant international instruments and processes to enhance policy coherence. These include biodiversity-related conventions (Convention on International Trade in Endangered Species, Ramsar Convention on Wetlands, Convention on Migrating Species, and World Heritage Convention), the UN Framework Convention on Climate Change (UNFCCC), and the Convention to Combat Desertification (CCD), as well as intergovernmental organizations, such as the Food and Agriculture Organization, UNESCO, the Intergovernmental Oceanographic Commission, and the World Trade Organization (WTO). There is a joint Web site for biodiversity-related conventions (see Further Reading, at the end of this article).

The CBD and the Ramsar Convention on Wetlands are implementing a joint work plan, including the River Basin Initiative. The secretariats of the CBD and the CCD are developing a joint work program. The COP has also called for greater cooperation with the UNFCCC on issues such as drylands, forest biodiversity, coral reefs, and incentive measures. The Intergovernmental Panel on Climate Change examines linkages between climate change and biodiversity and prepares scientific advice on the integration of biodiversity considerations into implementation of the UNFCCC and its Kyoto Protocol. The Secretariat is collaborating with UNESCO on the development of a global initiative on biodiversity education, training, and public awareness. The COP also emphasizes the importance of developing a common understanding of the relationship between the CBD and the WTO agreements, including the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS). It has called for further work with respect to the relationship between intellectual property rights (IPRs), relevant provisions of TRIPS, and the CBD, in particular those on technology transfer and on traditional knowledge and biological diversity. The COP has invited the WTO to explore the interrelationship between the CBD and the TRIPS. The COP has also sought to initiate cooperation with the World Intellectual Property Organization (WIPO) on the issue of IPRs arising from the implementation of the convention, such as those in access and benefit sharing and article 8(j) and related provisions. A memorandum of cooperation was signed with WIPO in 2003.

See also Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS); Convention on International Trade in Endangered Species (CITES); Global Environment Facility; multilateral environmental agreements; trade and the environment

FURTHER READING

Convention on Biodiversity Handbook. 2003. Montreal: CBD Secretariat and UNEP. Complete text of agreement found in this publication.

Global Biodiversity Outlook. Periodic report. Montreal: CBD Secretariat and UNEP. www.biodiv.org/GBO2. Periodic report on the state of biodiversity worldwide.

Hendrickx, Frederik, Veit Koester, and Christian Prip. 1993. "Convention on Biological Diversity: Access to Genetic Resources, A Legal Analysis." *Environmental Policy and Law* 23 (6): 250 58. One of the first studies to spell out necessary legal policy measures for achieving access and benefit sharing among users and providers of genetic resources.

Heywood, Vernon H, ed. 1995. *Global Biodiversity Assessment.* Cambridge: Cambridge University Press. An early and central UNEP empirical examination of biodiversity loss.

Kloppenburg, J. R. 1988. *First the Seed.* Cambridge: Cambridge University Press. A central theoretical and empirical analysis examining global interdependence in plant genetic resources for food and agriculture.

Laird, Sarah, and Rachel Wynberg. 2005. *The Commercial Use of Biodiversity: An Update on Current Trends in Demand for Access to Genetic Resources and Benefit Sharing, and Industry Perspectives on ABS Policy and Implementation.* UNEP/CBD/WGABS/4/INF/5. Nairobi: UNEP. A thorough empirical examination of

present and future needs for access to genetic resources within different sectors.

Millennium Ecosystem Assessment (MEA). 2005. *Ecosystems and Human Well Being: Biodiversity Synthesis.* Washington, DC: World Resources Institute. A comprehensive report and synthesis undertaken by more than 1,300 experts from 95 countries on the questions of how and why ecosystems change, and how this affects human well being.

Swanson, Timothy, ed. 1995. *Intellectual Property Rights and Biodiversity Conservation.* Cambridge: Cambridge University Press. A detailed, interdisciplinary analysis of economic and scientific rationales for biodiversity conservation, including clarification of the property rights issue.

Ten Kate, K., and Sarah A. Laird. 1999. *The Commercial Use of Biodiversity: Access to Genetic Resources and Benefit sharing.* London: Earthscan. The most comprehensive and practical guide to understanding the implications of and the implementation of the Convention on Biological Diversity.

Wilson, E. O., ed. 1988. *Biodiversity.* Washington, DC: National Academy Press. The earliest and most comprehensive book published on the issues of risk, value, and conservation of biodiversity, written in an accessible style.

Web sites for main biological diversity agreements and organizations are given below:

Convention on Biological Diversity. www.biodiv.org.

Global Environment Facility. www.gefweb.org.

The joint website for biodiversity related conventions. http://www.biodiv.org/convention/partners websites .asp.

G. KRISTIN ROSENDAL

■ Convention on International Trade in Endangered Species (CITES)

The Convention on International Trade in Endangered Species (CITES) is a mature international treaty that became effective in 1975. At the beginning of 2007, 169 sovereign states had ratified or acceded to the treaty.

Global trade in animals and plants is not measured by any one organization or individual; it occurs as $20 tourist items or as single blue fin tuna worth $40,000 for the Japanese sushi market, as well as reptile skin shipments worth more than $1 million. The total amount of trade in wild plants and animals is very difficult to determine, but one comprehensive estimate for the 1990s suggested an annual value of $15 billion, excluding timber and fisheries, which by themselves may have a value of $140 billion. Estimates of illegal trade in endangered species range between $5 billion and $10 billion per year. It is widely accepted that the value of illegal wildlife trade is second only to the illegal drug trade.

Approximately 5,000 species of animals and 28,000 species of plants are listed for some level of protection under CITES. Most well-known animals and plants are within the umbrella of the treaty: bears, elephants, whales, large cats, rhinos, sea horses, corals, orchids, and mahogany are but a sample of the species. The only group that is not well represented on the list is commercial fish from the open ocean. This absence has to do more with international politics than biological status.

The purpose of the treaty is to control the international movement of treaty-listed wild plants and animals, alive or dead, whole or parts thereof (specimens of species) in such a manner as to be assured that the pressures of international trade do not contribute to the endangerment of a species. This treaty does not deal with trade issues within a country or with the issues of habitat protection or preservation. CITES is implemented by way of permit requirements for the listed species. These permits are operative at the customs level as trade moves across national boundaries.

This is an active treaty regime; representatives of all the member states meet every two years to consider what adjustments need to be made to the list of protected species and what management and policy issues need to be addressed.

Listing of Species The treaty creates categories into which species of concern or at risk of extinction may be placed: appendix I and appendix II. Those on appendix I are species threatened with extinction and which are affected by international trade. Appendix II lists species that *may be* threatened with extinction and are part of international trade.

The placement of a species in appendix I or II is a group decision made at the conference of the parties (COP) and requires a two-thirds vote of the parties. Any state that disagrees with a listing decision may take a reservation on the species listing within 90 days of the vote. As of 2007, approximately 228 mammals and 146 bird species were listed on appendix I, and 369 mammal and 1,401 bird species on appendix II.

Since a consequence of listing a species on appendix I is that commercial trade in the species will be prohibited, proposals for placing new species on appendix I receive very close scrutiny by the delegates at the COP.

Permit Responsibilities Once a species is placed on either appendix I or II, then the protection of the treaty is triggered by the treaty requirement that each state must prohibit the transboundary movement of the species unless a permit has been issued by the relevant country. As there is no international police system to enforce the obligations of CITES, the treaty presumes that enforcement will be done at the national level. Indeed, a specific obligation of the treaty is for each state to adopt domestic legislation that will carry out the requirements of the treaty.

Appendix II species normally will be allowed in international trade so long as a CITES export permit has been issued by the state of export. Before issuing such a permit, the authorities of the state of export must make a key ecological finding: the scientific authority of the state of export must advise that such export will not be detrimental to the survival of that species. This is a particularly challenging requirement in developing countries where wildlife science for many species is nonexistent. If detrimental trade is allowed, knowingly or unknowingly, then the goal of the treaty is frustrated.

Appendix I species, those already identified as at risk of extinction, require two permits: an exporting permit and an importing permit. The criteria for the exporting permit are the same as with appendix II species. Note that the importing states must also make a finding that the purpose of the import will not be detrimental to the species. For the import permit the key requirement is that the management authority of the state of import must be satisfied that the specimen is not to be used for primarily commercial purposes. Although the determination of what is a primarily commercial purpose would seem to be straightforward in the main, on the margins it is a difficult question. Sport hunting trophies are generally not considered to have a commercial purpose even though they may have a market value. Importation of animals by zoos is likewise not prohibited by CITES custom even though the zoo will pay for the animal and generate money from its display.

The limitation on commercial use of appendix I species is controversial, in that some states argue that economic utilization of appendix I species would be useful for obtaining the resources and political motivation for the protection of the species for example, the sale of elephant ivory could raise funds to help protect elephants. Additionally some states are troubled by the fact that the importing county (primarily the developed countries) can block trade that the exporting country (primarily developing countries) believes is acceptable. These issues have been settled by the language of the treaty and it would take an amendment of the treaty to change the existing approach. Although an amendment process is provided for in the treaty, it is highly unlikely to occur.

Enforcement The enforcement problems faced by CITES are not just the limitations of the treaty language but also limitations within individual party states:

Lack of adequate domestic laws.

Lack of an adequate number of government employees and lack of pay and training for the employees that do exist.

Lack of scientific experts within a country, unwillingness to give power to them, and a lack of resources for the scientists that are present.

Lack of support from the police and courts for wildlife crime prosecutions and lack of serious punishment when a prosecution is successful.

Lack of a public education component.

Wild plants and animals are important to both local and international economic activity. The overconsumption of these resources is a critical issue in many countries. The CITES treaty provides a legal mechanism and obligation to control trade in endangered species and to protect these resources. In an environment of limited government resources, however, enforcement is often difficult.

See also multilateral environmental agreements; trade and the environment

FURTHER READING

CITES, Secretariat. The language of CITES, the list of species, and adopted management resolutions are available at www.cites.org, the web site of the treaty.

———. 2003. *CITES Handbook.* Geneva: Secretariat CITES. Contains a detailed explanation of all the aspects of the treaty.

Oldfield, Sara, ed. 2003. *The Trade in Wildlife: Regulation for Conservation.* London: Earthscan. Presents a series of essays that focus on the nature of wildlife trade and how different regulatory systems, including CITES, affect the species at issue.

Robinson, John G., and Kent H. Redford, eds. 1991. *Neotropical Wildlife Use and Conservation.* Chicago: University of Chicago Press. Contains a number of chapters discussing the economic use of wildlife; although older, the ideas and examples are still fully relevant.

Traffic. Factual information about the nature and consequences of trade can be found at www.traffic.org. Traffic is a science based organization that is a joint project of the World Wildlife Fund for Nature and the World Conservation Union.

DAVID FAVRE

■ convertibility

Convertibility refers to the ability of a currency to be freely transformed into foreign exchange. There are two types of convertibility: current account convertibility and capital account convertibility. The first term applies when the purpose of such transformation is to acquire foreign goods or services activities that are typically recorded in the current account. The second term is used when the purpose is the acquisition of foreign assets, such as foreign stocks and bonds activities that are typically recorded in the capital account.

Currency convertibility has direct consequences for international trade, capital mobility, and economic growth. When a country imposes restrictions on the conversion of domestic currency for foreign exchange, it necessarily impedes transactions between domestic residents or entities and foreign ones: most foreign counterparts will not accept a currency that is hard to convert into foreign exchange as a form of payment. Without current account convertibility, the cost of engaging in international transactions increases, leading to fewer transactions and to a lower level of trade. So important is current account convertibility that member nations of the International Monetary Fund (IMF) are required to fully adopt it under article VIII, sections 2, 3, and 4 of the Articles of Agreement.

Capital Account Convertibility and Growth A similar consensus, however, does not exist regarding capital account convertibility. The debate is more empirical than theoretical in nature. Theory tells us that in a perfect world, countries with full capital account convertibility can successfully exploit international capital markets to smooth out their consumption patterns over time; such countries can borrow from the rest of the world at a time when domestic consumption and investment are high relative to domestic income, and lend to the rest of the world when domestic consumption and investment are low relative to domestic income (Cole and Obstfeld 1991; Obstfeld and Rogoff 1995). Countries that have the ability to smooth out these income-expenditure patterns are theoretically better off than they would be otherwise. Without the ability to tap into international capital markets, a country would have to rely primarily on its own savings to fund domestic investment. Such a situation could limit future economic growth if domestic savings could not match the domestic level of investment demand.

According to the theory, countries that adopt and maintain full capital account convertibility should be

able to grow faster than countries that do not. Empirical studies, however, have not been able to settle this issue convincingly. Some researchers have found that convertibility has a positive effect on growth; others argue that it has either no effect or a negative effect.

During the late 1970s, many countries systematically removed restrictions on — that is, liberalized — their capital accounts. The experiences of these countries presented an opportunity for researchers to test the hypothesis that liberalization is associated with subsequent output growth. Although some researchers found empirical support for this hypothesis (e.g., Quinn 1997), others found positive effects only for high-income countries, not for emerging-market economies (Edwards 2001). More recent research has examined the microeconomic consequences of capital account restrictions to ascertain whether they affect different industrial sectors or even firms differently.

A case study of capital controls in Chile during the 1990s found that these restrictions changed the way firms chose to do their financing. Specifically, firms in Chile switched from relying primarily on short-term debt to relying on internal funds for financing investment spending (Gallego and Hernandez 2003). This change is what one would expect to observe if capital controls introduced a cost wedge between domestic sources of funds and international capital markets. Although the results suggest that capital controls were costly for at least some firms, they cannot be used to judge the desirability of imposing or removing controls. Such policy decisions must be based on their effects on society as a whole rather than on a particular sector of society.

Although some evidence suggests that convertibility influences growth at some stages of development, research also suggests that such influence is limited at best and negative at worst. For example, one study estimates standard growth regression equations and does not find that capital account convertibility is associated with long-run growth (Rodrik 1998). Another study finds that many European countries enjoyed impressive growth rates during the 1950–73 period not despite having capital account restrictions, but instead precisely because they had them (Wyplosz 1999).

Theory versus Reality Why does research find capital account restrictions to be beneficial for growth when basic economic theory tells us otherwise? The discrepancy appears to stem from the fact that the "real world" is more complicated than basic theoretical models make it out to be. In a world with asymmetric information (i.e., buyers and sellers do not have the same information) and other capital market distortions, the conditions for smoothly operating markets do not exist. Many researchers note that with free capital mobility — or full capital account convertibility — a country is exposed to the whims and fads of investors and traders who make trading decisions for all kinds of reasons, some of them rational but many of them seemingly irrational. When enough of them decide to pull their investments out of a country, the sudden demand for foreign exchange may trigger a currency crisis with devastating consequences for the banking sector and the rest of the economy. Work by Kaminsky and Reinhart (1999) offers empirical evidence substantiating this effect.

The lack of consensus in the empirical literature reflects the beneficial effects of full capital account convertibility for some countries but not others (Prasad et al. 2003). As a result, the literature has shifted its focus to ask whether countries must have a set of institutions or economic conditions before they adopt full capital account liberalization, and whether the sequence of capital account liberalization matters. For the first question, researchers have hypothesized that sound macroeconomic conditions — such as low and controlled fiscal deficits or low and controlled inflation — are crucial for countries considering full capital account liberalization (Sen 2007; Williamson, Griffith-Jones, and Gottschalk 2005). Other researchers believe that countries must implement a set of reliable institutions in the financial sector, such as imposing minimum information requirements, prudential supervision and regulation of the banking sector, and investor protection, when pondering the adoption of full current account convertibility (Chinn and Ito 2002; Arteta et al. 2001; Gelos and Wei 2002).

Research has also investigated whether the method and timing of liberalization (or the

sequencing of liberalization) matter as well. It is well known that short-term capital flows (such as portfolio investment) tend to be much more volatile than long-term flows (such as foreign direct investment). Given that this is the case, it is more prudent to liberalize long-term capital flows before liberalizing short-term ones. The idea is, of course, to reduce the country's exposure to what the literature has popularized as "hot money": short-term funds that enter and leave countries very suddenly and unexpectedly.

China and India are examples of countries that have considered adopting full capital account convertibility. Some commentators have argued that countries should always strive to adopt full capital account convertibility as soon as possible in order to reap all of its benefits (Forbes 2005). Others, however, argue that given the potentially disastrous consequences of implementing a liberalization program too rapidly, it is more prudent to go at it slowly and only when the country is "ready" (that is, when it has met all of the necessary preconditions). Given the theoretical complexities involved and the ambiguity of the empirical results, the advice that all researchers should probably offer to countries embarking on a full convertibility program is "proceed with caution."

See also asymmetric information; balance of payments; banking crisis; Bretton Woods system; capital controls; capital mobility; currency crisis; financial crisis; gold standard, international; hot money and sudden stops

FURTHER READING

Arteta, C., B. Eichengreen, and C. Wyplosz. 2001. "When Does Capital Account Liberalization Help More than It Hurts?" CEPR Discussion Paper No. 2910 (August). London: Centre for Economic Policy Research. Examines the evidence that capital account liberalization enhances growth and argues that correcting macroeconomic imbalances is a prerequisite for liberalizing the capital account.

Chinn, M., and H. Ito. 2002. "Capital Account Liberalization, Institutions, and Financial Development: Cross Country Evidence." NBER Working Paper 8967. Cambridge, MA: National Bureau of Economic Re

search. Investigates how capital controls affect the level of financial development.

Cole, H., and M. Obstfeld. 1991. "Commodity Trade and International Risk Sharing: How Much Do International Financial Markets Matter?" *Journal of Monetary Economics* 28: 3 24. Uses a general equilibrium model to quantify the social gains from having full capital account convertibility.

Edwards, S. 2001. "Capital Flows and Economic Performance: Are Emerging Economies Different?" NBER Working Paper 8076. Cambridge, MA: National Bureau of Economic Research. Studies the effect of capital mobility on economic growth.

Eichengreen, B. 2001. "Capital Account Liberalization: What Do the Cross Country Studies Show Us?" *World Bank Economic Review* 15: 341 66. Summarizes the results of many studies that investigate the role of capital account liberalization in economic performance.

Eichengreen B., M. Mussa, G. Dell'Ariccia, E. Detragiache, G. M. Milesi Ferretti, and A. Tweedie. 1999. "Liberalizing Capital Movements: Some Analytical Issues." *Economic Issues* 17 (February): 1 19. Summarizes the costs and benefits of having full capital account mobility, and touches on sequencing issues when countries decide to embark on a capital account liberalization program.

Forbes, Kristin J. 2005. "Capital Controls: Mud in the Wheels of Market Efficiency." *Cato Journal* 25 (1): 153 66. Argues that the use of capital controls brings microeconomic inefficiencies to capital markets, and hence that it may be prudent to eliminate all of them.

Gallego, F., and F. L. Hernandez. 2003. "Microeconomic Effects of Capital Controls: The Chilean Experience during the 1990s." *International Journal of Finance and Economics* 8: 225 53. Shows that Chile's experiment with capital controls resulted in an increase in the cost of capital for a sample of firms.

Gelos, R. G., and S. J. Wei. 2002. "Transparency and International Investor Behavior." NBER Working Paper 9260. Cambridge, MA: National Bureau of Economic Research. Finds that portfolio investment tends to be lower in countries that are less "transparent" that is, where it is hard to obtain clear and reliable financial market information.

Kaminsky, G., and C. Reinhart. 1999. "The Twin Crises: The Causes of Banking and Balance of Payments Pro

blems." *American Economic Review* 89 (3): 473 500. Studies the relationship between banking and financial crises.

Obstfeld, M., and K. Rogoff. 1995. "The Intertemporal Approach to the Current Account." In *Handbook of International Economics*. Vol. 3. Amsterdam: Elsevier/North Holland, 1731 99. Develops a comprehensive open economy model in a dynamic setting that can be used to measure the welfare benefits from having full capital account convertibility.

Obstfeld, M., and A. Taylor. 2002. "Globalization and Capital Markets." NBER Working Paper 8846. Cam bridge, MA: National Bureau of Economic Research. Documents and explains the evolution of world capital flows from the 19th century on.

Prasad, E., K. Rogoff, S. J. Wei, and M. A. Kose. 2003. "Effects of Financial Globalization on Developing Countries: Some Empirical Evidence." IMF Occasional Paper No. 220. Washington, DC: International Monetary Fund. Examines the empirical evidence be hind the claim that financial globalization promotes economic growth.

Quinn, D. P. 1997. "The Correlates of Changes in Inter national Financial Regulation." *American Political Sci ence Review* 91: 531 51. Investigates the political and economic variables that appear to be best associated with international financial regulation.

Rodrik, D. 1998. "Who Needs Capital Account Convert ibility?" In *Should the IMF Pursue Capital Account Convertibility? Essays in International Finance No. 207*, edited by P. Kenen. Princeton, NJ: Princeton University Press, 55 65. Describes how full capital account con vertibility may be more harmful than helpful to emerg ing market economies.

Sen, P. 2007. "Capital Inflows, Financial Repression, and Macroeconomic Policy in India since the Reforms." *Oxford Review of Economic Policy* 23 (2): 292 310. Provides a detailed analysis of the capital account liber alization process in India. The main conclusion is that, given the fragility of India's financial sector, full capital account convertibility for India may be premature.

Williamson, J., S. Griffith Jones, and R. Gottschalk. 2005. "Should Capital Controls Have a Place in the Future International Monetary System?" In *The Future of the International Monetary System*, edited by M. Uzan. Northampton, MA: Edward Elgar, 135 67. Provides a critical appraisal of the main arguments for and against capital controls and reports the main trends in the adoption and practice of capital controls across countries over the last 10 years.

Wyplosz, C. 2001. "Financial Restraints and Liberalization in Postwar Europe." In *Financial Liberalization: How Far? How Fast?* edited by G. Caprio, P. Honohan, and J. E. Stiglitz. New York: Cambridge University Press, 125 58. Looks at the evolution of financial repression and capital account liberalization in Europe.

CARLOS D. RAMIREZ

■ **copyrights**

See Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS); intellectual property rights

■ **corporate governance**

Corporate governance determines, among other things, how control over the firm's resources is allocated, and how control and a monitoring hierarchy are created within a firm. Corporate governance acts as a facilitator, enabling managers and shareholders to move toward the optimal governance structure within the firm. Even though policymakers are often concerned with finding an optimal legal, institutional, and regulatory framework to protect the interests of investors and other stakeholders, it is difficult, in practice, to establish a single model that meets the needs of all parties. Given the broad and vague definition of *corporate governance*, there can be situations where the importance of other stakeholders such as employees and creditors may be central to a corporate governance system. Naturally, a corporate governance regime does more than regulate the ownership and control arrangements inside the firm. A "good" corporate governance system, for example, provides rules and institutions that enforce internal ownership and control arrangements, but it also contains measures that protect other stakeholders from the opportunistic behavior of

insiders. Such a framework, which is mainly concerned with a company's investors but extends to wider interests, arguably provides a solid analytical basis on which a framework for understanding the key components of a good corporate governance system can be created.

What Is Corporate Governance and Why Does It Matter? The importance of corporate governance has become increasingly recognized for companies' and countries' prospects (Claessens 2006). As a consequence, regulators and policymakers in many countries diagnose weaknesses in their legal regimes and propose new arrangements said to foster higher firm performance, greater entrepreneurship, and better developed capital markets. Recent audit-based scandals in Europe and the United States have prompted investors to pay more attention to the institutions of corporate governance (Armour and McCahery 2006). Moreover, as allocation and monitoring of capital among competing purposes and investments have become more complex, financial instruments, corporate structures, and accounting have all become more sophisticated and often more risky. This has made monitoring of management by shareholders and other stakeholders more difficult, involving higher costs. The other side of the coin, of course, is that firms with good governance will benefit through higher market premiums (Gompers, Ishii, and Metrick 2003).

Corporate Governance Systems There are two main systems of corporate governance, which have significant differences: (1) the market-oriented corporate model, and (2) the relationship-based (or network-oriented) corporate system (Bratton and McCahery 1999; Roe 2003).

Market-oriented systems are characterized by dispersed equity holding through capital markets, a portfolio orientation among equity holders, and a broad delegation to management of discretion to operate the business. Two governance problems are said to result. The first is the shareholder management agency problem. Collective action problems prevent close monitoring of management performance by widely dispersed shareholders of small stakes. Imperfect performance incentives result

for managers, who may rationally sacrifice shareholder value to pursue their own agendas. Market systems address this management incentive problem with three corrective mechanisms: the hostile takeover, shareholder legal rights against management self-dealing, and the inclusion of outside monitors on boards of directors. The second productive disadvantage of the market system is a time horizon cost that stems from shareholders' tendency to rely on short-term performance numbers. This problem has been attributed to information asymmetries. Management has superior information respecting investment policy and the firm's prospects, but this information tends to be soft or proprietary and therefore cannot credibly be communicated to actors in trading markets. At the same time, market systems fail to provide clear-cut protections to managers who make firm-specific investments of human capital, a failure due in part to these systems' reliance on takeovers, proxy fights, and boardroom coups to control agency costs.

Market systems have countervailing advantages. Their shareholders can cheaply reduce their risks through diversification. Compared to shareholders in relationship-based systems, they receive higher rates of return on their investment. Market systems' deeper trading markets facilitate greater liquidity. These capital markets also facilitate corporate finance, providing management with greater flexibility on the type and source of new capital than do the markets in relationship-based systems. More generally, they provide an environment relatively more conducive to management entrepreneurship, as reflected in increased investment in new technologies.

Relationship-based systems are characterized by a majority or near majority of stock held in the hands of one, two, or a small group of large investors who hold blocks of shares. Like market systems, relationship-based systems leave management in charge of the business plan and operations. But large block investments imply a closer level of shareholder monitoring (Becht, Bolton, and Röell 2003). In addition, the coalescence of voting power in a small number of hands means earlier, cheaper intervention in the case of management failure. The other

primary benefit of relationship-based systems stems from the blockholders' access to information about operations. This decreased information asymmetry permits blockholders to invest more patiently. The longer shareholder time-horizon, in turn, frees management to invest for the longer term and creates a more secure management environment for firm-specific investment of human capital.

There are corresponding costs and benefits. Where the blockholder is a firm, internal agency costs can constrain its effectiveness as a monitor. Indeed, whatever the identity of the blockholder, its heightened oversight incentive does not appear to result in sharp oversight of management investment policy. Freedom to make long-term investments thus often means pursuit of growth in market share at the cost of suboptimal return on equity investment. Trading markets in relationship-based countries tend to be thinner and less transparent than in market-oriented countries. Meanwhile, the blockholders themselves forgo the benefits of diversification and, given the thin trading markets, the possibility of easy exit through sale. Finally, there is a shortage of loyalty. Blockholders, having sacrificed diversification and liquidity, extract a return in the form of private benefits through self-dealing or insider trading (Dyck and Zingales 2004). Legal regimes in network-oriented systems facilitate this quid pro quo with lax protection of minority shareholder rights and lax securities market regulation. The judiciary in network-oriented systems, unlike some market-oriented systems (which play a much more proactive role in shaping the actual contents of the corporate governance framework than in continental European jurisdictions), is more confined to interpreting the statutes and codes enacted by legislators (LaPorta et al. 1997).

Typology of Legal Strategies for Protecting Investors There is a wide collection of strategies for protecting shareholders and creditors from expropriation by managers and controlling shareholders. A basic division is between two categories: (1) regulatory strategies, through which the law mandates the terms of relationships among principals, agents, and the firm; and (2) governance strategies, through which the law channels the ongoing articulation of

the terms of corporate agencies (Kraakman et al. 2004). The strategies within each category are designed to provide different solutions to principal-agent problems ex post or ex ante.

On examination, it turns out that there are two categories of corporate law that pursue regulatory strategies. The strategies in the first category impose performance mandates on agents. These agent constraints apply on an ex ante basis by rule and on an ex post basis through open-ended standards of conduct. The strategies in the second category govern the terms of the principal's financial engagement with the firm. These affiliation terms apply on an ex ante basis when capital is transferred to the firm and on an ex post basis when capital is withdrawn or shares are sold. There are three other categories of corporate law measures that pursue governance strategies. The first category, appointment rights, concerns the ex ante selection of agents and their removal. The second, decision rights, concerns control over the terms of the firm's governing contracts and its business plan. Ex ante decision rights control the initiation and amendment of investments, divestitures, contracts, restructuring, and corporate legislation; ex post rights go to the ratification or veto of investments, divestures, contracts, restructuring, and legislation. The third set of governance strategies, agent incentives, concerns the incentives of the firm's agents. Ex ante these relate to the agents' qualifications as responsible fiduciaries; ex post they are related to the agents' financial rewards. From this analysis ten categories emerge, which are given in table 1.

A typology can be used to create indexes that capture the effectiveness of these five strategies to protect shareholders. The evidence shows that the strategy of giving shareholders strong legal rights in the appointment or removal of board directors has more measurable benefits for shareholders in common law countries such as the United Kingdom, which has tough rules on the removal of directors, than civil law regimes in continental Europe, where management has more discretion to address shareholder and creditor interests, and there are limited circumstances where a director can be removed. The evidence also points to potentially high benefits of

Table 1
Regulatory and governance strategies

	Regulatory strategies		Governance strategies		
	Agent constraints	Affiliation terms	Appointment rights	Decision rights	Agent incentives
Ex ante	rules	entry	selection	initiation	trusteeship
Ex post	standards	exit	removal	veto	reward

Source: Kraakman et al. (2004).

shareholder involvement in corporate decision-making, such as in the case of directors' remuneration. In the case of affiliation rights, the exit right of shareholders is addressed by corporate law strategies, developed albeit differently across Anglo-American jurisdictions, that effectively constrain management in frustrating a hostile bid and facilitate an exit right by ensuring shareholder approval of the bid. Changes in continental European takeover law to promote affiliation rights are still required, notwithstanding efforts by the European Commission to provide better constraints on target management in the Takeover Directive (McCahery and Renneboog 2003).

By comparison, the agent incentives strategies are often left to the contractual discretion of the board of directors. Codes of conduct set forth requirements in the area of board structure, the committee structure of the board, the role, monitoring and performance of the board and nonexecutive directors, and directors' remuneration. Codes of conduct and stock exchange rules draw up highly specific requirements concerning conflicts of interests and transactions where shareholder approval is necessary. According to a recent study that measures the board's independence from large shareholders, the legal strategies that deal with constraining the board and giving shareholders more authority are more effective in Anglo-American jurisdictions than civil law countries. Meanwhile, the empirical studies that focus on the reward strategy are voluminous (see, e.g., Jensen and Murphy 2004). This line of research reveals a number of problems in the design of appropriate criteria that link shareholder welfare to performance

criteria for directors and managers (Bebchuk and Fried 2004).

The above typology suggests a broad approach to corporate governance: (1) market value is the principal measure of shareholder interest; (2) other constituents should be protected by contract and outside regulation; (3) ultimate control should rest with the shareholders; (4) managers should be obligated to manage in the interests of the shareholders; and (5) noncontrolling shareholders should be strongly protected. These organizing factors, and the five corresponding regulatory strategies that constrain major shareholders and managers, provide a sufficient basis for evaluating corporate governance systems in differing jurisdictions and regions.

Good Corporate Governance In recent years, the increased focus on good corporate governance among academics, policymakers, and business leaders has given rise to investigations about how shareholders' rights and judicial enforcement vary across countries. This work has led to discussions which show that specific legal rules and practices are necessary to run a business and maintain corporate governance. Furthermore it shows that the best practices vary across countries (LaPorta et al. 1997). The impact of legal rules and institutions may vary, for example, depending on the judiciary's ability to quickly resolve disputes (Berglöf and Claessens 2004).

Similarly there is a growing awareness of how the variety of corporate governance arrangements and extralegal mechanisms are specific to firm type, industry sector, and sources of finance (Carlin and Mayer 2000). These findings suggest that a good

corporate governance system requires, inter alia, legal rules that promote a system of disclosure and accounting and audit systems that function effectively to provide investors with timely and accurate information to make effective investment decisions. The key message for lawmakers is that the agenda for governance reform is broad and challenging, but delivering in the key areas will enhance performance (Black, Jang, and Kim 2006).

The positive impact of corporate governance can be felt not only on firm performance but also on the level of financial market development. Countries with better property rights and strong investor protection tend to have higher-valued securities and deeper capital markets. Enhanced creditor-protection rights are also associated with deeper and more developed banking and capital markets. Recent theoretical and empirical work highlights that countries with better investor protection are also less vulnerable to external economic shocks. In the end, corporate governance can induce countries to adopt higher standards which can have a major impact on realizing their objectives.

See also infrastructure and foreign direct investment; multinational enterprises

FURTHER READING

Armour, John, and Joseph A. McCahery, eds. 2006. *After Enron: Reforming Company and Securities Law in Europe and the United States*. Oxford: Hart Publications. A wide ranging account of the recent corporate gover nance scandals and postscandal reform in Europe and the United States.

Bebchuk, Lucian, and Jesse Fried. 2004. *Pay without Per formance: The Unfulfilled Promise of Executive Compen sation*. Cambridge, MA: Harvard University Press. One of the most engaging and important discussions of the limitations of current executive pay practices in the United States, which provides a number of policy rec ommendations.

Becht, Marco, Patrick Bolton, and Ailsa Röell. 2003. "Corporate Governance and Control." In *The Hand book on the Economics of Finance*, edited by George Constantinides, Milton Harris, and Rene Stulz. Am sterdam: North Holland, 1 109. The leading review of the main features of corporate governance in both blockholder and market oriented regimes.

Berglöf, Erik, and Stijn Claessens. 2004. "Enforcement and Corporate Governance." World Bank Policy Re search Working Paper No. 3409. Washington, DC: World Bank. Available at http://www.ssrn.com/. An exploration of the key variables that help explain what constitutes an effective system of enforcement and gov ernance.

Black, Bernard, Hasung Jang, and Woochan Kim. 2006. "Does Corporate Governance Predict Firms' Market Values? Evidence from Korea." *Journal of Law, Eco nomics, and Organization* 22 (2): 366 413. One of the more recent studies that provides a sophisticated em pirical examination of the instrumental variables that affect the market value of companies.

Bratton, William W., and Joseph A. McCahery. 1999. "Comparative Corporate Governance and the Theory of the Firm: The Case against Global Cross Reference." *Columbia Journal of Transnational Law* 38 (2): 213 97. One of the earliest articles to evaluate the tradeoffs be tween the blockholder and market based system of governance.

Carlin, Wendy, and Colin Mayer. 2000. "How Do Fi nancial Systems Affect Economic Performance?" In *Corporate Governance, Theoretical and Empirical Per spectives*, edited by Xavier Vives. Cambridge: Cambridge University Press, 137 68. A sophisticated theoretical and empirical analysis of the governance needs of dif ferent types of firms over their life cycle.

Claessens, Stijn. 2006. "Corporate Governance and De velopment." *The World Bank Research Observer* 21 (1): 91 122. A important look at the role of corporate gov ernance in facilitating development and economic well being.

Dyck, Alexander, and Luigi Zingales. 2004. "Private Ben efits of Control: An International Comparision." *Journal of Finance* 59 (2): 537 600. The leading empirical study into the private benefits of control from a comparative standpoint.

Gompers, Paul, Joy Ishii, and Andrew Metrick. 2003. "Corporate Governance and Equity Prices." *Quarterly Journal of Economics* 118 (1): 107 55. An important empirical study of the way corporate governance provi sions in the United States impact firm value.

Jensen, Michael, and Kevin Murphy. 2004. "Remunera tion: Where We've Been, How We Got to Here, What Are the Problems, and How to Fix Them." ECGI Working Paper No 44/2004. Available at http://www .ecgi.org/wp. An interesting historical account of the system of executive compensation that provides a series of recommendation for reforming the system.

Kraakman, Reinier, Paul Davies, Henry Hansmann, Gerard Hertig, Klaus Hopt, Hideki Kanda, and Edward Rock. 2004. *The Anatomy of Corporate Law: A Comparative and Functional Approach.* Oxford: Oxford University Press. A lucid theoretical and comparative study of corporate law systems based on a typology of strategies that pro vides a way to compare the competing regimes.

LaPorta, Rafael, Florencio Lopez de Silanes, Andrei Shlei fer, and Robert Vishny. 1997. "Legal Determinants of External Finance." *Journal of Finance* 52 (3): 1131 50. An early study that looks at 49 countries, showing that countries with poorer investor protections have smaller and narrower capital markets.

McCahery, Joseph A., and Luc Renneboog. 2003. *The Economics of the Proposed Takeover Directive.* Brussels: Centre for European Policy Studies. An economic ac count of the development of the Takeover Directive in Europe.

Organisation of Economic Co operation and Develop ment (OECD).1999. *Principles of Corporate Governance.* Paris: OECD. The key principles for OECD and non OECD countries to evaluate and improve the legal and regulatory institutions for corporate governance within their countries and to provide guidance for other parties that play a role in the process of developing good cor porate governance.

Roe, Mark J. 2003. *Political Determinants of Corporate Governance: Political Context, Corporate Impact.* Oxford: Oxford University Press. An important account of the theory of the firm and the variation in governance rules across systems.

JOSEPH A. MCCAHERY

■ corruption

Corruption is commonly defined as the abuse of public office for private gain. It involves bribe taking and deviation from public duties by officials in var ious governmental spheres such as the grant of licenses and permits, implementation of public programs, and enforcement of government policies. This view of corruption is meant to be a working definition only and it does not capture the context-specific and multidimensional nature of corrupt acts.

The difficulties in arriving at a commonly ac cepted definition are many. Corruption has moral, legal, and sociological dimensions. It need not be confined to the public sector. Corruption need not involve bribes or monetary payments. Moreover, in certain societies the culture of gifts, networking, and reciprocal favors is quite common, and these acts are socially acceptable.

Corruption, in one form or other, is present in most countries in varying degrees. Many poor and devel oping economies, however, are perceived to be highly corrupt, and the recent attention in academic as well as policy circles tends to focus on the type of corrup tion prevalent in these countries. Many development economists view corruption as a major problem.

Corruption is a fairly new field of inquiry in economics, although the earlier literature on rent-seeking activities is related and has influenced re search in corruption. The phenomenon of corrup tion, however, is certainly very old. There are several accounts of governmental corruption in England, Italy, and China in the 18th and 19th centuries. References to bribery and the punishments for bribery can be found in many ancient sources, such as Babylon's Code of Hammurabi (22nd cen tury B.C.), Egypt's Edict of Harmhab (14th century B.C.), and Kautilya's Arthasastra (14th century B.C.).

Types of Corruption Many scholars make a dis tinction between two types of corruption: grand corruption and petty corruption (e.g., Rose-Acker man 1999). The latter category refers to bureaucratic corruption in which officials take bribes to grant fa vors in violation of their formal duties. Grand cor ruption refers to corruption by politicians and high-level decision-makers who accept payments from clients to influence policies or laws. This can be viewed as a form of state capture.

It is also useful to make a distinction between collusive and extortive corruption. Suppose an official is in charge of screening projects by granting approval only to socially desirable projects. Collusion occurs when the owner of an unqualified project manages to get approval by bribing the official. Extortion occurs when the official demands a bribe to approve qualified projects. The granting of permits and licenses by bureaucrats with near monopoly power resembles this type of corruption. Corruption can also take the form of embezzlement and appropriation of public funds. This is the more dominant form of corruption in many public education, health, and antipoverty programs.

Measuring Corruption Irrespective of their legal status, corrupt acts are always held in secrecy, and it is difficult to unearth systematic information on corruption. To a large degree research on corruption has suffered on this account. Early discussions of corruption tended to rely on journalistic accounts in newspapers and reports, on anecdotal evidence, or in a few cases, on prosecutorial evidence. Recently, there have been several attempts to construct more systematic measures of corruption.

Many authors have used perception measures of corruption. These measures are based on perceptions by various types of respondents such as local and multinational businessmen, country experts, and citizens. Two common measures are the corruption perception index (CPI) of Transparency International and the corruption control index (CCI) of the World Bank Institute. Transparency International has been publishing these indexes on an annual basis, and they now cover more than one hundred countries (Transparency International 2006). The CCI is more recent and it is available for selected years (Kaufmann et al. 2004). These indexes are based on many different polls conducted by several independent organizations. These polls use different methods and different sets of questionnaires about the business environment and corruption.

More recently, there have been attempts to use experience-based measures of corruption. The World Bank conducted a Business Environment and Enterprise Performance Survey (BEEPS) in 1999, in which selected firms in 26 East European and central Asian countries were asked various questions about the extent of bribe payments and the reasons for making such payments. In general, these surveys take care to elicit bribe-related information from the firms. Firms are seldom asked about their own payments; rather they are encouraged to talk about the average payment made by a similar firm in the industry.

Similar surveys have been conducted in some other countries to elicit information about corruption-related experiences of the general public. Despite their imprecise nature, these surveys convey useful information about corruption in different branches (police, judiciary, public services) of the government. There have been recent attempts to collect such information through field experiments.

In addition to the perception-based and experience-based measures, economists have also looked at several indirect measures. For example, by tracking government expenditure from its origin to the destination, one can get an estimate of the extent of leakage of public funds. This leakage may be viewed as a measure of corruption.

Causes Recent research into the causes of corruption focuses on the role of factors such as an oversized public sector, poor quality of regulatory institutions, the lack of democratic governance, low wages of public officials, and lack of economic competition. Researchers have used country-level perception measures to explore these causal links. Although there is disagreement about the nature and strength of these links, one can identify certain common features of the countries perceived to be highly corrupt.

In general, low levels of development (measured in terms of per capita income) and low human capital (measured in terms of years of schooling) are associated with high levels of perceived corruption. These two can be viewed as structural features, which could change only in the long run. But there are several other features that reflect the policy choices made by these countries.

For example, most of these countries also have a highly regulated business environment with a plethora of rules and bureaucratic procedures. A study by

Djankov et al. (2002) shows that in a cross section of countries the minimum official time required for a business startup varies from 2 business days in Canada and Australia to a high of 152 days in Madagascar. The association between the number of days required to start a legal business and the extent of perceived corruption is positive and robust. The presence of the rules allows bureaucrats to exercise monopoly power and extract bribes from private business.

A related feature is the absence of economic competition. The regulatory environment tends to discourage entry of new firms and helps to maintain the profitability of the few existing firms. These existing firms, in turn, transfer part of their profits as bribes to the politicians and bureaucrats. The concentration of economic power and wealth is more pronounced in smaller economies where economic activities rely on abundant natural resources or primary exports. The initially unequal distribution of assets (land, natural resources) leads to concentration of wealth in the hands of a few who use the corrupt system to increase their own wealth at the cost of others.

In many of these countries salaries of public officials are substantially lower than salaries in the private sector. This creates an inducement for the officials to seek illegal bribe income. Since in most cases the punishment for bribery involves loss of job and future income, low salaries also reduce the cost of engaging in corruption.

Other features such as lack of strong democratic traditions and lower level of press freedom are associated with higher levels of corruption. Among the democratic countries the forms of the government and the nature of electoral competition also seem to be associated with different levels of corruption. A presidential form of government is likely to be less corrupt than a prime ministerial form because of greater accountability, but this result crucially depends on the nature of checks and balances in place. Similarly, among electoral rules, a proportional representation system of voting tends to be associated with higher corruption. In a proportional representation system voters do not directly elect any particular political candidate, reducing the ability for voters to punish corrupt officials.

It is possible that noneconomic variables such as culture, values, religion, and geography also play roles in contributing to the levels of corruption. In a widely cited paper, Triesman (2000) finds support for the view that countries with histories of British rule and Protestant traditions tend to be less corrupt. Though such a result is by no means robust, it is a reminder that cultural determinants of corruption cannot be ignored altogether.

Consequences Some scholars argue that corruption can enhance efficiency. In the presence of a rigid bureaucracy and ill-planned regulations, corruption can facilitate speedy implementation of productive investment proposals. According to this view, corruption could grease the wheels of commerce and promote growth. There is very little support for this view in contemporary theoretical or empirical research, however. At best, corruption could result in a partial and temporary efficiency gain. Officials who benefit from such grease money would always have a tendency to increase red tape for their own benefit. Recent research has shown that corruption leads to inefficient allocation of resources with adverse effects on growth, investment, and foreign direct investment.

Mauro (1995) initiated one popular strand of empirical research on corruption by examining the link between corruption and growth in a cross section of countries. Using the CPI as an explanatory variable, he observed that growth and investment rates are adversely affected by corruption levels. Later studies tend to confirm this broad relationship, though the exact nature depends on several region-specific variables. One possible exception to this relationship is the experience of many high-growth Asian countries that are perceived to be highly corrupt. Though there is not enough evidence to suggest that corruption has promoted growth in these countries, it suggests that the corruption-growth link is quite complex.

Various studies have claimed that corrupt countries tend to have high inflation rates, greater military expenditures, and less public investment in education and health. The direction of causation is far from established, but the possibility that corruption

distorts governmental spending programs cannot be ignored.

Corruption adversely affects inequality and poverty also. Poverty is affected because corruption tends to lower growth, affects the (pro-poor or otherwise) orientation of public spending, and undermines the implementation of many of the antipoverty programs. This issue has generated concerns among national governments, nongovernmental organizations, and international organizations such as the United Nations, the World Bank, and the European Union. Empirical research in this area is somewhat limited, however.

At a more micro level, corruption undermines enforcement policies in a variety of fields such as tax collection, environmental regulation, and policing. Hence a higher level of corruption would lead to lower tax revenues, poor environmental standards, and a deteriorating law and order situation.

Anticorruption Strategies Views on anticorruption strategies differ considerably, ranging from mild tolerance to advocacy of draconian measures. Even though corruption is perceived to be a problem associated with low levels of development, it is unlikely that corruption will fade out on its own with rising income levels in these countries. First, corruption might constrain the growth process itself. Second, corruption has a tendency to spread and persist across time. Third, apart from generating various economic inefficiencies mentioned earlier, corruption undermines people's faith in democracy and other human development objectives. Hence curbing corruption is both a goal in itself and an instrument for achieving other goals.

Anticorruption strategies involve design of suitable incentives and organizational structures and, at a more general level, improvements in institutional quality. To a large extent, a person's decision to be corrupt is based on calculations of associated costs and benefits. Hence many have proposed incentives measures that seek to reward honest behavior and punish the corrupt. This, of course, requires better monitoring and information gathering and a quick and fair judicial process. Such measures have proved to be successful in some of the recent tax enforcement

reforms exercises, such as those carried out by the Philippines, Singapore, and Brazil. Studies based on field experiments also support the view that better monitoring is effective.

Additionally, one can reduce the scope of corruption by increasing transparency, reducing discretion, and encouraging competition where appropriate. This involves measures such as simplification of laws and easier access to information by the public. There must also be ways to generate and transmit credible and hard information about governmental decisions where there is potential for corruption. A recent study shows how a newspaper campaign about the amount of funds allocated to schools in certain areas helped stakeholders (parents) to hold the disbursement officials accountable and reduce the leakage of funds. The role of local bodies, activists, and an independent press is quite vital in this respect. Last, bureaucratic and other forms of corruption are unlikely to thrive without corrupt politicians. As pointed out earlier, electoral competition and forms of democratic governance influence the scope and extent of corruption by political representatives. Hence a successful anticorruption strategy may require reforms in the political and electoral process as well.

See also aid, international; political economy of trade policy; smuggling

FURTHER READING

Bardhan, Pranab. 1997. "The Economics of Corruption in Less Developed Countries: A Review of Issues." *Journal of Economic Literature* 35 (3): 1320 46. An excellent review of the issues and some of the early literature.

Djankov, Simeon, Rafael La Porta, F. Lopez de Silanes, and Andrei Shleifer. 2002. "The Regulation of Entry." *Quarterly Journal of Economics* 117 (1): 1 37. One of the first empirical exercises to capture the nature and impact of bureaucratic red tape across countries.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2004. "Governance Matters III: Governance Indicators for 1996, 1998, 2000 and 2002." *World Bank Economic Review* 18 (2): 253 87. Perception indexes for various aspects of governance including corruption in a large set of countries.

Klitgaard, Robert. 1988. *Controlling Corruption*. Berkeley: University of California Press. A popular book containing various case studies and proposals for anticorruption reforms.

Mauro, Paolo. 1995. "Corruption and Growth." *Quarterly Journal of Economics* 110 (3): 681 712. An important contribution evaluating the impact of corruption on investment and growth in a cross section of countries.

Mishra, A. 2005. *Economics of Corruption*. Oxford: Oxford University Press. A reader containing a survey of the literature and several important academic contributions in the field.

Rose Ackerman, Susan. 1999. *Corruption and Government: Causes, Consequences and Reforms*. Cambridge: Cambridge University Press. An excellent commentary on various aspects of governmental corruption by one of the pioneers in the field.

———. 2006. *International Handbook on the Economics of Corruption*. Cheltenham, UK: Edward Elgar. An excellent collection of handbook style survey and research articles by leading scholars in the field.

Svensson, Jakob. 2005. "Eight Questions about Corruption." *Journal of Economic Perspectives* 19 (3): 19 42. An excellent discussion of the key issues in corruption.

Transparency International. 2006. *Global Corruption Report 2006*. London: Pluto Press. An annual report of Transparency International discussing its CPI, country analysis, and recent empirical research on corruption.

Triesman, Daniel. 2000. "The Causes of Corruption: a Cross National Study." *Journal of Public Economics* 76 (3): 399 457. An extensive cross country analysis of determinants of corruption.

AJIT MISHRA

■ corruption and foreign direct investment

See infrastructure and foreign direct investment

■ countervailing duties

Countervailing duties (CVDs) are tariffs imposed by a country to offset a foreign subsidy. The General Agreement on Tariffs and Trade (GATT) has permitted the imposition of CVDs since 1947 to offset "actionable" subsidies if the government determines that these subsidized imports are causing or threatening to cause injury to the domestic industry. Actionable subsidies are subsidies that are not explicitly prohibited under the GATT but are subject to challenge if they cause adverse effects on another country.

CVDs are similar to anti-dumping duties, the tariffs imposed to protect domestic industries from products imported at unfairly low prices, in that both seek to counteract the "unfair" trade practices of a foreign trade partner. Anti-dumping duties are much more prevalent than CVDs in the world today, however. Between 1995 and 2005, World Trade Organization (WTO) members imposed 1,804 anti-dumping duties compared to just 112 CVDs. As a result, CVDs have provoked significantly less research and debate than anti-dumping duties.

Some economists emphasize that importing countries should welcome subsidized imports rather than punish the exporting country through the use of CVDs because foreign subsidies lower import prices and increase domestic welfare at the expense of the foreign government. Other economists and lawyers stress that CVDs can improve welfare by eliminating economic distortions caused by subsidies and discouraging the future use of these subsidies. It is unclear whether current CVD regulations are solely used to offset foreign subsidies; some empirical evidence suggests that CVDs are often geared more toward imposing additional trade protection rather than eliminating economic distortions that arise from foreign subsidies. In order to understand the current debate over CVDs, it is important to understand the economic arguments for and against the use of CVDs, the evolution of CVD regulations and their worldwide use, and how future international trade negotiations could alter the use of CVDs.

Economic Consequences of CVD Laws Economists generally agree that CVDs can improve global economic welfare, at least in principal. Consider, for example, Canadian lumber exports to the United States. Trade theory suggests that Canadian subsidies would lower the price of Canadian lumber exports to

the United States. The United States would benefit from these lower prices as long as it is a net importer of lumber. In other words, U.S. consumers of lumber would gain more from the lower, subsidized prices than U.S. lumber producers would lose from lower prices and the loss of market share to Canadian producers. In this situation, the imposition of a CVD would increase the import price, restore the welfare of U.S. lumber consumers and producers to their pre-subsidy levels, and transfer money from the Canadian government to the U.S. government.

The imposition of the CVD causes total U.S. welfare to fall in this scenario because the gain to the U.S. government and U.S. lumber producers does not offset the loss that U.S. consumers experience from the higher prices. Even in this situation, however, the imposition of the CVD could result in an increase in global economic welfare. If the subsidy allows Canadian producers to export lumber to the United States even though U.S. producers harvest lumber more efficiently, then it creates an economic distortion. A CVD that offsets this subsidy should eliminate the unnatural advantage and the distortion, thus improving total economic welfare. By imposing the CVD, however, the U.S. government improves world welfare at the expense of a net decrease in U.S. economic welfare.

Since the late 1980s, most economists have agreed that under certain conditions CVDs can improve domestic welfare. Assume that there are two firms serving the world market for aircraft, one in the United States and a second in the European Union. Both firms, taking the output of their foreign competitor into consideration, choose output levels to maximize their profits. European production subsidies will artificially lower the cost of production, causing the European producer to increase its output. The U.S. producer will react by cutting its production level to try to maintain a higher price. The subsidy lowers the world price of aircraft, although not by the full amount of the subsidy as in the lumber example. It also increases European profits and lowers U.S. profits. The production subsidy can increase total European welfare as long as the gain to European consumers and producers offsets the cost

to the government. In contrast, the subsidy may decrease U.S. welfare if the loss of profits to the domestic aircraft producer is greater than the gain to U.S. consumers.

Using theoretical models, both Dixit (1988) and Collie (1991) show that the optimal response of the U.S. government in a situation like this is to retaliate with a CVD that partially offsets the subsidy. The CVD increases European production costs, thereby increasing the price of aircraft and transferring some of the profits of the European producer to the U.S. government and the U.S. producer. The CVD can increase U.S. welfare in this case because the gain to the U.S. producer and government will offset the loss to the U.S. consumer.

Collie also shows that a foreign country should be deterred from subsidizing exports when a domestic country uses CVDs. This theoretical prediction implies that the threat of the imposition of CVDs should be enough to eliminate subsidies. The continued use of subsidies and CVDs, however, reveals that this is not the case. Qiu (1995) develops a model to show that the coexistence of subsidies and CVD regulations can be explained by things such as lengthy delays in the imposition, and restrictions on the level, of CVDs.

Countervailing Duties from 1890 to the Present In 1890, the United States enacted the first CVD law, which applied only to certain grades of sugar. Belgium enacted the first CVD law targeting all subsidized imports in 1892, and the United States expanded its CVD coverage to all imports in 1897. By 1921, nine additional countries had passed CVD legislation.

The 1947 GATT formally sanctioned the use of CVDs to counteract foreign subsidies. The CVD portion of the 1947 GATT was largely based on the U.S. law at the time, with one important exception. The GATT allowed for the imposition of CVDs only if subsidized imports were causing or threatening to cause injury to a domestic industry. The United States was granted a grandfather provision and thus did not require proof of injury until 1974. International rules governing the imposition of CVDs remained largely unchanged until the 1979 Tokyo

Round GATT Agreement on Subsidies and Countervailing Measures and the 1994 revision of this agreement that resulted from the Uruguay Round of trade negotiations.

Under the version of the agreement finalized in 1994, governments may initiate a CVD investigation at the written request of the domestic industry. Prior to imposing a CVD, the government must determine that the imports from the country under investigation have (1) benefited from a prohibited or actionable subsidy, and (2) the subsidized imports have caused or threatened to cause "material" injury to the domestic industry. If the government makes a final determination that subsidized imports have injured the domestic industry, it may impose a CVD up to the calculated amount of the subsidy per unit of imported product.

In general, the agreement does not detail how countries should measure either the amount of the subsidy or the degree of injury to the domestic industry. The agreement does specify, however, that the amount of the subsidy should be consistent with some generally accepted guidelines, such as:

- The difference between the amount paid by the firm on a government loan or a government-guaranteed loan and the amount the firm would pay on a comparable commercial loan obtained in the private market; and
- The difference between the market price of a good and the price paid to the firm by the government in the purchase of goods, or the difference between the market price of a good and the price paid by the firm to the government in the provision of goods.

Injury determinations should be based on an examination of the volume of subsidized imports, the impact of these subsidized imports on domestic output and prices, and the subsequent impact of the imports on the domestic industry. In conducting this examination, the government should evaluate all economic factors associated with the industry, including the actual and potential decline in output, market share, profits, and capacity utilization, as well as other factors that may be causing injury to the

domestic industry. No CVDs may be imposed if the government determines that the amount of the subsidy is less than 1 percent of the import price or the volume of subsidized imports is negligible.

Investigations must be concluded within 18 months of their initiation. CVDs should remain in force only for as long as necessary to counteract the subsidies causing injury. The duties must be eliminated after five years unless the government determines during a sunset review that revocation of the duties would be likely to lead to a recurrence of injury caused by the subsidies.

Countries have a great deal of latitude in the way they conduct CVD investigations. Most countries have either a bifurcated approach, in which the injury and subsidy investigations are conducted by two different governmental entities, or a single-track approach, in which both investigations are conducted by the same entity. For example, in the United States, the International Trade Administration of the Department of Commerce determines the level of subsidies, while the International Trade Commission determines whether the subsidies have caused injury to the domestic industry. In contrast, the Trade Directorate of the European Commission makes both determinations in the European Union.

In 1956, 20 countries had CVD legislation, although only 8 actually used the legislation. The United States, which has historically made much greater use of CVDs than other countries, imposed only 62 CVDs between 1897 and 1959. Under the Tokyo Round Agreement, however, the use of CVDs by a small subset of countries rose dramatically. For example, the United States imposed CVDs in more than 90 cases between 1980 and 1988 alone. Between 1989 and 1993, 7 signatories to the GATT initiated 150 CVD investigations, of which the United States accounted for slightly more than 45 percent.

Since the completion of the Uruguay Round Agreement, more countries have imposed CVD measures, although its use is still highly concentrated among a few countries. Seventeen WTO members initiated 182 CVD investigations between 1995 and 2005. Of these, 61.5 percent, or 112 investigations,

eventually resulted in the imposition of CVDs. The United States and the European Union initiated 64 percent of all CVD investigations between 1995 and 2005. Half of all CVD investigations during this time period were filed against exports from India, Korea, Italy, Indonesia, and the European Union. Metals, particularly steel products, accounted for nearly 40 percent of all CVD petitions; other leading industries targeted for CVD actions include food products and footwear.

Some economists argue that many of these CVDs were used primarily to protect domestic industries from targeted imports, whether subsidized or not. For example, Marvel and Ray (1996) find that U.S. CVD petitions tend to be made against multiple nations simultaneously and are closely linked with anti-dumping petitions. Moreover, most CVD investigations define any financial contribution to the foreign firm as a subsidy, regardless of whether it lowers the foreign export price and creates an economic distortion. Based on these characteristics, Marvel and Ray conclude that CVDs are used primarily to provide the domestic industry with protection from imports rather than to counteract subsidies.

Other researchers have found empirical evidence that political pressure may explain the imposition of some CVDs. As summarized in Blonigen and Prusa (2003), there is a large literature that tests the economic and political determinants of the injury decisions in both CVD and anti-dumping investigations. Most research in this area has analyzed U.S. investigation outcomes and found that although economic factors significantly influence outcomes, political pressure also positively influences the likelihood that a CVD or anti-dumping investigation will result in the imposition of duties.

The Future of CVD Laws Stiglitz (1997) suggests that CVD laws can be beneficial as long as they are not used to protect domestic industries from import surges. He and others suggest that regulations be revised to include a higher standard of proof for establishing the existence of subsidies. In 2001, the members of the WTO agreed to "clarify and improve" CVD regulations during the Doha Round of

trade negotiations, although the clarifications will not necessarily resolve the problems economists have with CVD laws. Negotiators developed a long list of proposals to consider, including specific changes to regulations governing the calculation of the amount of the subsidy and the injury determinations.

In summary, a countervailing duty is a tool that countries can use to offset the unfair use of production subsidies by foreign governments. The use of CVDs is relatively rare today compared to other trade remedy tools such as anti-dumping duties. There is no clear consensus among economists on the impact of CVDs on the economy. Although traditional international trade theory predicts that the imposition of a CVD will increase global welfare at the expense of the imposing country, more recent theoretical models have shown that CVDs can improve domestic welfare under certain conditions. Empirical evidence suggests that CVDs may be geared more toward imposing additional trade protection rather than offsetting foreign subsidies; therefore some economists have urged the WTO to revise current regulations governing the use of CVDs.

See also anti-dumping; Doha Round; fair trade; General Agreement on Tariffs and Trade (GATT); new trade theory; subsidies and financial incentives to foreign direct investment; Uruguay Round; World Trade Organization

FURTHER READING

Blonigen, Bruce A., and Thomas J. Prusa. 2003. "Anti dumping." In *Handbook of International Trade*, edited by E. K. Choi and J. Harrigan. Oxford, UK: Blackwell, 251 84 . Includes a review of the literature testing the economic and political determinants of injury decisions in CVD and anti dumping investigations.

Collie, David. 1991. "Export Subsidies and Countervailing Tariffs." *Journal of International Economics* 31: 309 24. Theoretical study that suggests that CVDs should deter subsidization.

Congressional Budget Office. 1994. *How the GATT Affects U.S. Antidumping and Countervailing Duty Policy.* Wash ington, DC: CBO. A thorough review of the history of GATT and U.S. CVD and anti dumping regulations.

Dixit, A. K. 1988. "Antidumping and Countervailing Duties under Oligopoly." *European Economic Review*

32: 55 68. Theoretical study that proves that the optimal response of a country to a subsidy is a CVD.

Marvel, Howard P., and Edward John Ray. 1996. "Countervailing Duties." *Economic Journal* 105: 1576 93. A statistical analysis of the use of CVDs in the United States.

Qiu, Larry D. 1995. "Why Can't Countervailing Duties Deter Export Subsidization?" *Journal of International Economics* 39: 249 72. A theoretical study of the inability of CVDs to deter subsidization.

Stiglitz, Joseph E. 1997. "Dumping on Free Trade: The U.S. Import Trade Laws." *Southern Economic Journal* 64 (2): 402 24. A review of U.S. CVD laws before and after the Uruguay Round, and the policy objectives of these laws.

Viner, Jacob. 1923. *Dumping: A Problem in International Trade*. Chicago: University of Chicago Press. One of the earliest works to justify and describe anti dumping and CVD laws.

KARA M. REYNOLDS

■ **crisis prevention, management, and resolution**

See bail-ins; bailouts; currency crisis; expenditure changing and expenditure switching; International Monetary Fund (IMF)

■ **currency basket regime**

See band, basket, and crawl (BBC)

■ **currency board arrangement (CBA)**

A currency board is a monetary arrangement based on two simple rules. First, the exchange rate between the domestic currency and an appropriately chosen foreign currency is fixed. Ideally, the fixed exchange rate is written into law in order to signal a long-term commitment. Second, there is full convertibility between the domestic monetary base and the foreign anchor currency. The currency board stands ready to exchange domestic into foreign currency, and vice versa, at the fixed rate, on demand, and without limit.

These two defining features have immediate and important implications. Under a currency board arrangement (CBA), every note or coin of the domestic currency in circulation is backed by foreign currency. Therefore, the currency board needs to hold a large amount of liquid foreign exchange reserves at least 100 percent of the monetary base, and perhaps more in order to provide a cushion against potential capital losses on the reserves. Domestic assets (also known as "domestic credit") on the balance sheet of the currency board should be zero or, more realistically, should be held constant. There should be no scope for creating money by expanding domestic credit. In practical terms, this prohibits the monetary authority from lending to the domestic government or commercial banks. In other words, the currency board is not allowed to monetize budget deficits or to act as a lender of last resort to troubled commercial banks. Therefore, in order to succeed, a CBA must go hand in hand with fiscal discipline and a healthy banking system.

On the continuum of exchange rate regimes, a CBA falls between a conventional fixed exchange rate and an outright monetary union. Under a conventional peg, there is only partial coverage of the monetary base by foreign reserves and there is no long-term commitment to the level of the exchange rate. Under a monetary union, a country gives up its currency altogether. A CBA is a more credible arrangement than a conventional peg but less credible than a monetary union.

Under a CBA, the adjustment mechanism for achieving external equilibrium is completely automatic. The monetary authority plays a passive role; it simply stands ready to exchange domestic notes and coins into foreign ones, or vice versa, as necessary. The money supply is endogenous and determined solely by market forces. If a country faces a balance of payments deficit (perhaps because private capital inflows are insufficient to cover the current account deficit), the monetary authority will face a loss of foreign exchange reserves. Under the operating rules of the currency board, the money supply will con-

tract. In asset markets, this will push domestic interest rates up, which will attract greater private capital inflows into the country. In goods markets, the falling money supply will exert deflationary pressure, and prices and wages will tend to fall over time. This will make domestic tradable goods relatively cheaper in world markets, and the current account deficit may shrink. High interest rates and the real depreciation caused by deflation will stifle domestic aggregate demand, particularly the demand for imports, thus contributing further to the current account improvement.

This should sound familiar. It is very similar to the adjustment mechanisms operating under the classical gold standard. In fact, some economists have argued that the gold standard was a special case of a CBA, with a commodity (gold) replacing the foreign anchor currency. Given the key role of price and wage adjustment in the mechanism just described, a successful currency board needs flexible goods and labor markets.

History The history of currency boards spans approximately 150 years. There have been around 80 CBAs throughout the world. During the colonial era, various dominions of the British Empire operated more than 70 different CBAs. Currency boards combined centralized control by the colonial center with the retention of domestic monies in the periphery. They were designed to facilitate trade and financial flows within the British Empire. Currency boards fell into intellectual disfavor and were gradually replaced by conventional central banks after the demise of colonialism in the 1950s and 1960s. Central banks better fitted the political drive for national independence. The transition also reflected the period's intellectual climate, which was infused with a sunny optimism about the effectiveness of discretionary government policies.

Currency boards came back into fashion beginning in the early 1990s, due to a collapse of faith in discretionary monetary policy, especially in developing countries. A cursory look at some countries that have operated currency boards in the 1990s and early 2000s (Hong Kong, Argentina, Estonia, Lithuania, Bulgaria, and Bosnia and Herzegovina) shows

that CBAs have been successful in ending hyperinflation, facilitating the transition to a market economy or to national independence, assisting postwar reconstruction, and restoring stability in an international financial center plagued by political uncertainty and banking crises. CBAs were also proposed in the aftermath of macroeconomic crises in Indonesia, Russia, Brazil, and Turkey, although none of these proposals ultimately came to fruition. Modern CBAs have been subject to occasional speculative attacks   Hong Kong in 1997, and Argentina in 1995 and 2001. In general, exits from CBAs have been uneventful, with the single spectacular exception of Argentina in early 2002.

Most real-world CBAs have reserved some degree of discretionary powers, and thus deviate from the definition of an orthodox currency board. The ratio of foreign exchange reserves relative to the monetary base may be allowed to dip below 100 percent. This coverage ratio may also be allowed to go above 110 115 percent in order to sterilize capital inflows, finance a lender-of-last-resort function, or smooth fluctuations in domestic interest rates. The monetary authorities in almost all contemporary currency boards have reserved the right to change the required reserve ratio for commercial banks, which might thwart the operation of the automatic adjustment mechanism described earlier. (The monetary authority regains some degree of control over the broader money supply.) Almost all modern CBAs have volatile domestic assets on their balance sheets.

Argentina's CBA was the one with the most loopholes, which probably contributed to its ultimate demise. First, Argentina's CBA was allowed to lend to the government. Second, throughout 2001 the long-term commitment to the fixed exchange rate was undermined by switching the peg from the U.S. dollar to a currency basket, and by instituting a complicated scheme of export subsidies and import surcharges, which amounted to a sneak devaluation.

Debating the Pros and Cons Professional economists have advanced various arguments about the advantages and disadvantages of CBAs. Perhaps the single biggest advantage of a currency board is its simplicity and transparency. It may enable a

government to commit credibly to monetary discipline at the expense of flexibility. In essence, through the currency board, the monetary authority can import the anti-inflationary credibility of the central bank issuing the anchor currency, typically the U.S. Federal Reserve or the European Central Bank. The domestic monetary authority becomes a mere warehouse for foreign cash.

A credible CBA reduces currency and default risk, which leads to lower domestic interest rates. It can boost the development of long-term financial markets, which are sorely lacking in many developing countries. By reducing transaction costs, the fixed exchange rate can promote trade and foreign direct investment. Finally, by establishing monetary discipline, a currency board can serve as a catalyst for a broad range of other reforms.

A serious drawback of currency boards is that they make the adopting country more vulnerable to external shocks, particularly terms-of-trade shocks, large capital inflows or outflows, or shocks emanating from the country issuing the anchor currency. Arguably, high interest rates in the United States and the strength of the dollar against the euro and the Japanese yen contributed to the demise of Argentina's CBA, in addition to lack of fiscal discipline within the country, especially at the provincial level.

As Argentina's experience with deflation in 1999 2001 also demonstrates, currency boards (or indeed fixed exchange rates in general) are probably not a good monetary arrangement for relatively closed economies. The less open the economy is, the larger the real depreciation necessary to eliminate a given balance of payments deficit. Under a fixed exchange rate, real depreciation can happen only through a fall in domestic prices. Such price adjustments can be slow and painful.

One reason to introduce a CBA is to arrest high inflation, but currency boards can generate overvalued real exchange rates and large current account deficits. Real appreciation results from the combination of a fixed nominal exchange rate and persistent domestic inflation. The introduction of a currency board can also be burdensome and time-consuming. In order to make the arrangement credible, policymakers must build political consensus, rewrite laws, and reorganize institutions.

Critics of CBAs have pointed out that counter-cyclical monetary policy is impossible under a CBA, and therefore output, employment, and prices would be more volatile under such an arrangement. Proponents of CBAs have countered that ending discretionary monetary policy is the intended purpose of such institutional arrangements: discretionary monetary policies in developing countries have been the prime source of macroeconomic instability, and currency boards have helped to restore macroeconomic stability by imposing a monetary straitjacket on the government. Monetary policy is an extension of fiscal policy   printing money is a government's revenue source of last resort. Therefore, a currency board can be beneficial in imposing a hard budget constraint on the country's treasury.

Opponents have noted that currency boards either abolish or sharply curtail the capacity of the monetary authority to serve as a lender of last resort to the banking system. Proponents counter that a lender-of-last-resort function is still possible under a CBA through assistance from the treasury, a deposit insurance scheme, or contingent credit lines from abroad. Bailing out domestic banks involves redistribution of income, and a currency board removes a nontransparent (and therefore politically cheap) tool of income redistribution. By making bank bailouts more politically costly and therefore less likely, a currency board limits the moral hazard problem of the banking system (the problem of a party insulated from risk engaging in riskier behavior). In essence, it imposes a hard budget constraint on domestic financial institutions.

Currency boards are unable, by definition, to create money by expanding domestic credit. Therefore, they cannot earn extra seigniorage through discretionary money creation. Currency boards still make profits on the difference between interest earned on their assets (highly liquid foreign securities) and their operating costs. Seigniorage revenues are somewhat lower under currency boards than under conventional central banks, however, both because CBAs cannot expand domestic credit and because foreign reserves may pay a somewhat lower

interest rate than domestic assets. Of course, many economists would count the reduction in seigniorage as a *benefit* of CBAs, given that it is an inefficient and nontransparent way of raising revenue.

The Empirical Record Although theoretical arguments have dominated the debates about currency boards, little empirical work has been undertaken on the subject. The most authoritative empirical analysis available (Ghosh, Gulde, and Wolf 2000) finds that currency boards are associated with lower inflation than either floating or conventional fixed exchange rate regimes. This result is highly robust. Currency boards also appear to be associated with higher rates of gross domestic product (GDP) growth, although the reasons for this are not clear. It may be that countries with better overall economic policies self-select in choosing to establish a currency board. It may be that most countries introduce a currency board following a severe macroeconomic crisis, and the better growth performance observed in the first few years of the new arrangement reflects a postcrisis "rebound effect." At the very least, currency boards have not been associated with *lower* GDP growth rates, although there is some evidence that output tends to be more volatile. Finally, CBAs appear to be associated empirically with lower money supply growth rates, smaller budget deficits, and better export performance.

In conclusion, a CBA offers significant benefits but also has serious drawbacks. Although it is neither a quick fix nor a panacea for all economic ills, in some cases it can deliver monetary discipline and low inflation.

See also convertibility; currency substitution and dollarization; discipline; exchange rate regimes; expenditure changing and expenditure switching; foreign exchange intervention; gold standard, international; impossible trinity; international reserves; lender of last resort; money supply; quantity theory of money; reserve currency; seigniorage; sterilization

FURTHER READING
Ghosh, Atish, Anne Marie Gulde, and Holger Wolf. 2000. "Currency Boards: More Than a Quick Fix?" *Economic Policy: A European Forum* 31: 269 321. A comprehen

sive and sober look at the history and the pros and cons of CBAs. Contains the best analysis of the empirical record of currency boards.

Hanke, Steve, and Kurt Schuler. 1994. *Currency Boards for Developing Countries: A Handbook*. San Francisco: ICS Press. A treatise on the defining features, history, and benefits of currency boards.

Roubini, Nouriel. 1998. "The Case against Currency Boards: Debunking 10 Myths about the Benefits of Currency Boards." Downloadable from http://www.rgemonitor.com/. A succinct and passionate summary of the cons of currency boards.

Schwartz, Anna. 1993. "Currency Boards: Their Past, Present, and Possible Future Role." *Carnegie Rochester Conference Series on Public Policy* 39: 147 88.

Williamson, John. 1995. *What Role for Currency Boards?* Washington, DC: Institute for International Economics. Schwartz and Williamson each offer thoughtful and thorough analyses.

SLAVI SLAVOV

■ currency competition

Traditionally, *currency competition* refers to competition between privately issued monies or between privately issued and government-issued monies for use as means of payments. This entry will focus on such competition. Competition between government-issued monies the pound sterling and the euro, for example is discussed in the entry on multiple currencies.

Types of Money To understand the ideas related to currency competition it is useful to define a few terms. First, *commodity money* is money that has some intrinsic value. The value is embedded in the commodity that makes the coin. Gold and silver coins are examples of commodity monies. In principle, it is possible to melt such coins and the resulting metal would sell for a value approximately equal to that of the coins. In contrast, *fiat money* is intrinsically worthless. The paper on which a $20 bill is printed is worth very little. The value of such a bill comes from the belief that other people will accept it in exchange for goods or services. *Inside money* is a claim on the

assets of the issuer of that money. It is a liability of the issuer. Traveler's checks are an example of inside money, as the issuer must, by law, repay such a check. Another example is bank deposits. In contrast, *outside money* is not a claim on its issuer. An issuer of outside money does not have a legal obligation to exchange its notes against some goods or services. Government-issued notes are examples of outside money. Central banks are not required to back the notes they issue with any kind of commodity.

Hayek and Currency Competition The economist Friedrich Hayek introduced the idea of currency competition in 1976. At that time, a number of countries were experiencing high inflation. The collapse of the Bretton Woods system of fixed exchange rates in 1971 had removed a constraint on the conduct of monetary policy for many monetary authorities. It also removed any remaining link between currencies and a metallic standard (gold in this case). Freed from these constraints, many central banks increased their money supply too quickly, spurring inflation.

Although there was broad agreement that inflation was undesirable, it was not clear how to give monetary authorities incentives to maintain the value of their currencies. Hayek reasoned that the lack of incentives came from the monopoly that monetary authorities held over the issuance of money. Competition from private issuers could force these authorities to maintain the value of their currencies. Indeed, if a money lost too much of its value, consumers and businesses would have the option to use another money. Hayek proposed to give banks the authority to issue notes that would compete with government-issued currency.

The interest generated by Hayek's proposal led a number of economists to study historical episodes during which banks were allowed to issue notes. Such episodes typically involved commodity money, however. Since the value of commodity money is intrinsic, there is not much scope for an issuer of commodity money to modify the value of its currency.

Economists have also been interested in competing inside monies. Such monies are typically denoted in some outside money, however. For example,

traveler's checks are often denominated in U.S. dollars. For this reason, it does not appear that inside money can affect the incentives of the issuer of outside money in the way Hayek intended.

What Hayek specifically had in mind was competition between *outside fiat* monies. He proposed to allow banks to issue intrinsically worthless notes that would not be liabilities of these institutions.

Private Issue of Outside Money For the kind of competition that Hayek proposed to occur, private institutions must be able to issue outside money. A number of economists have argued that a time inconsistency problem makes private issuance impossible. These economists point out that if a private institution could issue valued fiat outside money, profit maximization would induce the institution to issue money up to the point where the marginal benefit of an extra unit of money equaled its marginal cost. That is, since the cost of producing an additional unit of fiat money is negligible, the institution would issue money up to the point where it was worthless. Moreover, the issuing institution would have no obligation to redeem the note for something of value since it would be issuing outside money. Anticipating this outcome, nobody would want to hold the money in the first place.

Private issuance of fiat outside money may be possible provided agents hold certain expectations. The key is to find a way to prevent issuers from increasing their supply of notes to the point where the value of such notes is zero. Suppose everybody believes that the notes of a private issuer have value if the number of notes issued is less than some number. This threshold number is common knowledge. If the number of notes issued exceeds the threshold, then everyone believes that these notes are worthless. A serial number on a note indicates the quantity of notes issued previously.

Consider someone who is offered a note with a serial number that is greater than the threshold level above which notes are no longer valued. This person believes that nobody will be willing to accept the note in exchange for goods or services in the future. Hence, this person will not accept the note, and the belief that the note has no value is self-fulfilling. This

kind of belief limits the ability of the issuer to print too many notes, which makes private issuance possible.

A Concept for the Future Although private issue of outside fiat currencies is possible in theory, there are few examples of such currencies. Some currencies used in online multiplayer games may be the best examples of fiat outside money. These currencies are almost costless to produce and are not backed by any asset. Moreover, online auction sites, such as eBay, provide a platform on which such currencies trade for U.S. dollars in a way that is similar to a currency exchange market. Although we are still far from Hayek's vision of currency competition, technological progress and new information technologies could make it happen.

See also discipline; dollar standard; dominant currency; money supply; multiple currencies; quantity theory of money; reserve currency; seigniorage; time inconsistency problem; vehicle currency

FURTHER READING

Martin, Antoine, and Stacey L. Schreft. 2006. "Currency Competition: A Partial Vindication of Hayek." *Journal of Monetary Economics* 53: 2085 2111. Provides a proof of the possibility of circulation of private outside money in the context of a formal model and a number of ad ditional references.

ANTOINE MARTIN

■ currency convertibility
See convertibility

■ currency crisis
Currency crises are among the most dramatic events in global financial markets. They generally involve large outflows of funds from currencies that investors fear may devalue or sharply depreciate. When a government lowers the level at which a currency is pegged to other currencies this is referred to as a devaluation, while a fall in the value of a market determined exchange rate is called depreciation. To keep their currencies from depreciating in the face of exchange market pressure, governments must run down their foreign currency reserves, borrow from abroad, hike interest rates, and/or impose capital controls to slow down outflows. If they are successful, the speculation dies down, at least for a while, and the currency value is maintained. More often, exchange rates are eventually forced to adjust. Although the vast majority of currency crises involve downward pressure on currencies, sometime there are strong market pressures for surplus countries to let their currencies appreciate; a vivid example is China beginning in the early 2000s. Crises can vary in their effects as well as in their causes, but they have in common a high degree of exchange market pressure.

International speculators are widely blamed for generating crises, especially by government officials eager to shift the blame from themselves. In reality, however, much of the pressure usually comes from "normal" domestic and international firms that are trying to protect themselves against losses rather than to generate speculative gains. Even the speculators are more often the messengers that there are problems than the independent cause of currency crises. Speculative attacks occur when international businesses and investors anticipate a change in the value of the currency, thus adding speculative outflows to the underlying balance of payments deficit.

Crises, though always uncomfortable, can sometimes have good effects, stimulating policy reforms and promoting economic recovery. In other cases, however, especially in developing countries with weak financial systems and considerable debt denominated in foreign currency, the effects of crises can include large recessions and widespread suffering. This is particularly likely when a currency crisis either leads to or was generated by a domestic banking crisis a so-called "twin crisis."

In general, the underlying causes of a currency crisis are inconsistencies between a country's exchange rate policy and its domestic economic policies so that the currency is either over- or undervalued relative to the currencies of other countries. Economists have developed a variety of theoretical models

to explain currency crises, and most of these models can be classified into three types. First-generation models consider cases where the government is either unable or unwilling to correct inconsistencies between its exchange rate and other domestic policy goals. As these become more serious, a crisis eventually becomes inevitable. Second-generation crisis models analyze cases where these inconsistencies place the economy in a "zone of vulnerability," making a crisis possible but not inevitable. In these cases, the potential for a crisis outbreak depends critically on market perceptions of the government's willingness and ability to take corrective action. Finally, newer crisis models have focused on financial sector weaknesses and the role of politics in the emergence of currency crises.

First-Generation Crises: Fundamental Disequilibriums The classic theoretical explanations for currency crises — the so-called first-generation models — focus on fundamental economic disequilibriums such as large government budget deficits. In these models, governments are assumed to pursue fiscal and monetary policies that are inconsistent with maintaining their fixed or slowly adjusting pegged exchange rate regimes. The resulting balance-of-payments deficits are financed by running down foreign reserves. When the level of reserves falls to a certain threshold, there is a sudden balance-of-payments crisis. Capital flight by domestic residents escalates, even if the country has borrowed little from abroad. This leads to a loss of reserves and forces the government to devalue its currency or float the exchange rate. In these models, expansionary policies and the resulting bad economic fundamentals push the economy into crisis.

The particular causes of the fundamental disequilibriums in past crises have been varied. They often result from a lack of fiscal discipline, but first-generation-type currency crises have also been caused by budget deficits in the center country of a pegged exchange rate system (the center country is the nation to whose currency another country's currency is pegged) or external developments such as export slumps. These problems generally coincide with weak governments facing strong political pressures

that prevent them from addressing the emerging disequilibriums by cutting spending or raising taxes to keep monetary expansion under control. Very often, this inability to balance the budget results in high rates of inflation. When economic problems emerge, markets initially often give governments the benefit of the doubt, especially when they have large stores of international reserves. But as imbalances continue, market participants frequently realize well before governments do that the problems are not temporary. The capital outflows that result from this realization involve not only currency speculators but also businesspeople wanting to hedge against the risk of exchange rate adjustments. These flows in turn tend to increase payments imbalances further and eventually lead to a crisis and a forced adjustment.

Examples of first-generation-style crises include the currency crises of Latin American countries in the post–Bretton Woods period, such as Mexico in 1976 and Brazil on several occasions. In these cases, the government financed large deficits by creating more money, resulting in inflation. Combined with pegged or slowly adjusting exchange rates, such high levels of inflation led to progressive overvaluation of exchange rates and generated numerous currency crises. Inflation does not need to be in triple or quadruple digits in order to generate currency crises, nor are inflation-driven crises a threat only to developing countries, as demonstrated by Italy's crisis in 1992. As a member of the European Monetary System (EMS), Italy had pegged its exchange rate to the deutschmark. Although its inflation rate was far below triple digits, it was still well above the inflation rates of most of its partners in the EMS, and in particular higher than that of the center country, Germany. The resulting disparity eventually led to a currency crisis in 1992. Large budget deficits can at times cause currency problems even if they do not lead to monetary accommodation and rising inflation; one example of this is the 2001 crisis in Argentina. As part of a "shock therapy" program designed to break out of a cycle of high and rising inflation, Argentina had adopted a currency board in 1991 that took money creation out of the hands of

the government by fixing the domestic currency to a foreign anchor currency, and maintaining full convertibility between domestic currency and the foreign anchor currency. This proved to be highly successful in bringing inflation under control but fiscal profligacy continued. As external developments contributed to a worsening of the situation, the result was a combined debt and currency crisis that ended in abandonment of the currency board and default on the government's foreign debt.

The collapse of the Bretton Woods system in the 1970s provides an example of an exchange rate system failing due to budget deficits in the center country (in this case the United States). For domestic political reasons, the U.S. government delayed tax increases required to pay for the large increases in expenditures associated with the Vietnam War. Seeking to prevent interest rates from escalating, the Federal Reserve Board financed a sizable portion of the resulting budget deficits with monetary accommodation. The consequent overheating of the U.S. economy led to increasing balance-of-payments deficits and ultimately to the currency crisis that was the final straw leading to the widespread abandonment of the Bretton Woods regime of adjustably pegged exchange rates.

Another type of first-generation crisis that need not be directly caused by bad domestic policies involves export slumps. Generally, monetary and fiscal policies that are sustainable under good circumstances become unsustainable when circumstances take a turn for the worse. In a country whose government's budget depends on export revenues, a sharp decrease in those revenues can cause a similarly sharp change in investor expectations of that country's ability to meet its debt-servicing obligations. A clear example of this is the 1998 Russian ruble crisis. By the mid-1990s Russia had tamed the runaway inflation that followed the breakup of the Soviet Union and its economic prospects had substantially improved. The domestic coalition favoring economic reform and stability was fragile, however. In the aftermath of the 1997 Asian crisis, oil prices fell to half of their precrisis levels due to decreased demand for oil in Asia. As a result, the Yeltsin government,

dependent as it was on oil revenues to fund its fiscal programs, found its treasury rapidly being emptied. Initially investors' belief that Russia was "too big to fail" and their perception that the Russian government was attempting to alleviate the situation by negotiating a loan package from the International Monetary Fund (IMF) and seeking to implement dramatic fiscal reforms caused most investors to give Russia the benefit of the doubt. After the IMF negotiations failed and the Russian legislature, the Duma, removed the most critical components of Yeltsin's proposed fiscal reform plan, however, it became clear that Russia was not going to be able to adjust its policy enough to avoid a fiscal crisis. In August 1998, there was a severe speculative attack on the ruble, and the Russian government announced that it would no longer support the crawling peg and would default on its foreign-held debt. The remaining investors immediately and frantically sought to divest themselves of the Russian liabilities and the ruble plummeted.

Second-Generation Crises: The Role of Expectations A surprising feature of a number of the currency crises of the 1990s was that they hit countries whose macroeconomic fundamentals were not particularly bad. First-generation models were unable to explain, much less predict, many of these crises. In response, a "second generation" of crisis models was developed that focused on investors' expectations and governments' conflicting policy objectives and predicted that speculative attacks could occur when a country's fundamentals were merely in an intermediate or vulnerable zone. Although these models retained the assumption that speculators would not attack countries with good fundamentals, they showed that once a country finds itself in a vulnerable zone, a change in the private sector's expectations about the future course of government policy can trigger a second-generation crisis.

In contrast to the passive role of policymakers in first-generation models, governments in second-generation models are able to take corrective measures, though they may prefer not to. For many governments, exchange rate stability is only one

objective among many. Governments often make a trade-off between maintaining a pegged exchange rate and achieving other policy goals (such as low unemployment) that gain political salience when there is an economic downturn. In such situations speculators can lose confidence in the government's commitment to the exchange rate peg and can decide suddenly to attack the currency. The cyclical state of the economy can thus become an important factor in these models, since the tight monetary policies necessary to defend a currency are politically more costly when the economy is in recession than when it is booming.

Speculators are also collectively in a position to influence the future course of policy in second-generation models. The greater the proportion of speculators who expect the government to defend the peg at the expense of a worsening recession, the smaller the capital outflows and the more feasible a defense of the exchange rate peg. On the other hand, if most speculators bet against a defense, capital outflows will be larger and the costs of a defense will be higher. Since these outflows themselves can be detrimental to a government's ability to defend the pegged exchange rate, market expectations can thus prove to be self-fulfilling. Such speculation is not irrational, nor is it in itself destabilizing in the sense of going contrary to the fundamentals. It is, however, sometimes thought of as destabilizing in the sense that it can be the proximate cause of a crisis that may not necessarily have been inevitable.

In formal second-generation models, shifts in investor expectations from optimistic to pessimistic are treated as arbitrary. The resulting outcomes are therefore often referred to as "sunspot equilibria." This can give a misleading impression, however, since in the typical cases of major speculative attacks on currencies, certain events trigger the shift from a "good" to a "bad" equilibrium. In the past, such triggers have included the assassination of a presidential candidate (Mexico in 1994) and crises elsewhere that lead to a reevaluation of underemphasized fundamentals (the wake-up call concerning financial sector weakness that was a major aspect of the contagion in Asia in 1997).

An example of what many experts consider to be a second-generation crisis is the 1992 93 crisis in the EMS, particularly the speculative attacks on the French franc. The proximate cause of this crisis was the reunification of Germany, which led to a large increase in German government expenditures. As German policymakers were unwilling to offset this higher spending with higher taxes, a large budget deficit emerged. In response, the Bundesbank (Germany's central bank) raised interest rates. Since Germany was the EMS center country, the logic of the currency system required all other EMS members to tighten monetary policy as well, even though this ran counter to the requirements of their macroeconomic situation at the time. As we have already discussed, this generated crises that led to the breakdown of the EMS system. The crises in Italy and, arguably, the United Kingdom were quite consistent with first-generation models since both the lira and the pound appeared to be overvalued. The 1993 attack on the French franc was not as easy to explain, however, since some of France's fundamentals were even stronger than Germany's.

What proved to be more important was the lack of willingness of the individual EMS countries to adjust to the mutual payments imbalances that had emerged. Here Germany held the cards. Even so the EMS member states put considerable pressure on the Bundesbank to lower interest rates, but the independent German Bundesbank refused to back off its tight monetary policies. When it became known that Germany would also not continue providing short-term financing to maintain the pegged rates of the EMS, the only alternatives left for the other EMS members were depreciation or a substantial tightening of their own macroeconomic policies. Given the state of the economy in many of these countries, international financial market participants therefore increasingly questioned policymakers' willingness to implement policies that would further slow growth and increase unemployment. Eventually, financial markets launched speculative attacks, forcing governments to choose between unpalatable options: abandoning the peg, which they had worked hard to maintain, or facing a deepening recession. In the case

of France, for example, the authorities withstood several bouts of speculative pressure before finally devaluing the franc in 1993.

The fact that France's fundamentals had not been particularly bad led many commentators, especially French officials, to conclude that France had been the innocent victim of destabilizing speculation. France was indeed an innocent victim in this case, but of geopolitical developments in Germany and its own commitment to the European pegged-rate system, not of capricious speculators. This illustrates that crises do not always fit neatly within one type of crisis model or another. Some economists consider France's crisis to be a classic second-generation example, but others argue that it was more of a first-generation crisis because of a fundamental disequilibrium generated by Germany's huge budget deficit. The crisis certainly had elements emphasized in both types of models, as did the 1997 Asian crisis, to which we now turn.

Financial Sector Weakness Most of the financial world and many economists were shocked in 1997 by the occurrence and the severity of the Asian financial crises. A crisis in the Thai exchange market had been at least partially anticipated, as there had been concerns about the possible overvaluation of the baht for some time. But the spread of the crisis to other countries was almost entirely unanticipated. Most if not all of the affected countries had enjoyed low inflation and robust economic growth for quite some time and appeared by and large to have strong enough economic fundamentals to preclude the type of crises predicted by first- and second-generation crisis models. In order to explain these events, economists developed what is sometimes referred to as the "third generation" of crisis models, which place a greater emphasis on the micro-foundations of currency crises. Among these refinements are the considerations of financial sector weakness, moral hazard, contagion, and stock as well as flow disequilibriums.

Traditional macroeconomic models paid little attention to the financial sector, and the crises of the 1990s demonstrated that this was a major mistake. The discovery of serious weaknesses in financial

sectors can generate major changes in international capital flows, especially where countries have weak international liquidity positions (i.e., high short-term foreign debt relative to international reserves). Such a situation can quickly turn into a run on the currency without requiring any outright speculation; the scrabble to cover open positions could be sufficient. Indeed, many of the flows of funds during the Asian crisis were of this risk-covering nature.

One of the factors that can contribute to financial sector weakness is a high level of moral hazard, referring in the currency crisis context to the propensity to lend, borrow, or invest in enterprises under circumstances that would usually be considered excessively risky. In some cases, lending institutions or borrowers may have had either explicit or implicit guarantees that the government would cover any losses should the loan or investment fail to return a profit. In this way the contingent liability burden is shifted from the lender, investor, or borrower to the government, creating the situation described by economist Paul Krugman (1996) as a bet of "heads, I win; tails, the taxpayer loses." If this practice is sufficiently prevalent, even a minor shock to the economy could result in a dramatic increase in government financing to prop up banks whose nonperforming loans have suddenly rendered them insolvent.

Many of the other factors described earlier, most notably financial sector weakness, also played key roles in precipitating the Asian crises of 1997. Although public finances in all of the crisis countries were relatively strong by conventional measures, domestic financial sectors were in bad shape. Both the high levels of nonperforming loans and heavy unhedged foreign borrowing were generated at least in part by problems of moral hazard. The combination of moral hazard and weak risk management and regulatory systems had led to many ill-advised loans, so that in a number of countries the financial sectors were suffering from serious solvency and liquidity problems. While good statistics were not generally available to reveal these problems, international funds continued to pour into these emerging markets because of their apparently strong

macroeconomic statistics and as a result of capital account liberalization.

In July 1997, the already overvalued Thai baht came under speculative attack when the severity of Thailand's problems with nonperforming loans in the financial sector became apparent. Highlighting Thailand's financial sector problems, this crisis served as a wake-up call to international financial markets, alerting them that Thailand's unhedged foreign borrowing was much riskier than many had believed. This in turn prompted foreign investors to reassess the vulnerability of their investments to financial sector risks. Thailand was not the only country found to have serious problems in the financial sector, and investors' assessments of some countries previously perceived to have well-aligned exchange rates, such as Korea and Taiwan, changed (those exchange rates were now seen as overvalued), prompting a run for the exits. Although financial sector problems were not the only contributing factors (as mentioned, the Thai baht was overvalued and political instability in Indonesia contributed importantly to the length and depth of the crisis there), the Asian crisis did highlight that crises could occur even when data on traditional economic fundamentals indicated a strong economic situation.

The Political Economy of Currency Crises Recent research has also recognized that political considerations can have an important influence on crises through a number of channels. Since economic policies are generally determined through the political process (one exception being when monetary policy is administered by an effectively independent central bank), political developments can have a strong influence on expectations about future economic policies. Thus political instability and expectations about the election of new administrations can have an important effect on capital flows even before there is any actual change in policy. Political considerations are also one of the major causes of the development of inconsistencies among policies that lead to fundamental disequilibriums. Because the costs of corrective action often show up faster than the benefits, short-run political pressures can lead to delays in undertaking needed policy adjustments,

especially if elections are approaching (see Willett 2007).

One key example of how political considerations can play a role in the outbreak of crisis is the 1999 crisis in Brazil. Past fiscal excesses had left Brazil with a large public debt burden, which was substantial but manageable so long as interest rates remained moderately low. In 1999, polls increasingly indicated that leftist presidential candidate Lula da Silva and another leftist politician were likely to prevail in the impending elections. Lula's past as a trade unionist raised doubts among market participants over whether Brazil's sound economic policies would be continued after the election. Although a return to hyperinflation was not likely, the high debt levels meant that even a moderate loosening of fiscal policy might be sufficient to force default. As a consequence, a crisis of confidence and a speculative attack on the real ensued. It is important to note that this crisis was precipitated not by actually implemented policies, but rather on the evaluation of expected future policy.

The 1994 Mexican crisis provides another example of the importance of political events in understanding crises. The country had successfully brought down inflation from triple- to single-digit levels, but the combination of strong capital inflows and a slow rate of depreciation designed to limit domestic wage increases had resulted in a large current account deficit and a substantially appreciated real exchange rate. Thus although Mexico's domestic economic fundamentals were strong, the economy was vulnerable to a drop in capital inflows, and as with the Brazilian crisis to severe shocks to political stability. In this case, the shock was the assassination of the leading opposition (and proreform) presidential candidate, which prompted rapid reassessment on the part of investors as to whether necessary reforms in Mexico would be implemented, along with fears of a more volatile political climate. The combination of monetary tightening in the United States, temporary preelection loosening of fiscal policy in Mexico, and the emergence of domestic political instability brought the dreaded sharp fall in inflows and the crisis was on.

Political scientist Andrew MacIntyre (1999) has argued that Thailand and Indonesia illustrate ways in which different types of governments can exacerbate their vulnerability to crises. Thailand's political system with its large coalition governments tended to produce great policy stability (or paralysis) due to the difficulties in making policy changes. Intracoalitional politics caused budget and financial regulation reforms to be delayed, further weakening confidence in the Thai financial sector and generating uncertainty about the government's capacity to act once the crisis occurred. In contrast, Indonesia's political system was highly centralized and imposed little constraint on executive action. This led to its own set of problems by opening the way for erratic policy behavior. When crisis spread from Thailand to Indonesia in late July 1997, the Suharto government responded decisively and preemptively by widening the band within which the rupiah was allowed to fluctuate from 8 percent to 12 percent. As the rupiah continued to slide, Suharto appeared to be taking appropriate action by agreeing to a series of dramatic reforms as part of an IMF assistance program. Appearances can be deceiving, however, and in this case they certainly were, as Suharto quietly reneged on many of these reforms and as members of his government and family openly undermined others. It is not surprising that investor confidence in the Indonesian government plummeted, not because investors doubted whether Suharto *could* make the necessary reforms but instead because they doubted that he *would*. Where Thailand's case can be characterized as inaction in the face of a looming crisis, Indonesia's can be characterized as action but of the wrong kind.

Crisis Prevention and Management There is widespread agreement that it is usually far easier to prevent crises than to fix them once they occur. Unfortunately, prevention is not easy to accomplish. Although analysis has often been able to identify sound reasons for crises after the fact, these reasons are not always easy to identify beforehand. The perception of crises as arising from the unjustified behavior of fickle financial markets doubtlessly stems at least in part from this dilemma. Significant advances have been made in the development of economic "early warning systems" for crises, though these tend to detect occasions when a country is vulnerable to a crisis rather than predict precisely when a crisis will occur.

Even when the threat of a crisis is unmistakable, however, countries often fail to take preventive actions in time. Sadly, it often takes the actual outbreak of crises to prompt major policy adjustments, as governments often find it politically difficult to take the actions needed to reduce vulnerability. Consequently, governments are often observed to initially respond to potential crises by addressing capital flows rather than macroeconomic fundamentals by running down international currency reserves, increasing domestic interest rates to make the country a more attractive capital destination, or implementing restrictions on the flow of capital into or out of the country. When the country is facing a temporary liquidity crunch, such measures may be all that is required (see the 1997 case of Hong Kong, where some of these measures were successfully used to avoid the most serious potential repercussions of the crisis). If, however, the country faces a fundamental disequilibrium (such as is described in first-generation crisis models), such measures can only delay, not prevent, crises; borrowing and running down reserves cannot indefinitely offset the effects of a fundamental disequilibrium. In such cases, reserve cushions can only give countries the option of adjusting more gradually. Second- and third-generation models, however, provide insights into ways that strong reserve positions can help avoid crises. When countries are in vulnerable zones, shocks that would generate a run on the currency of a country with high levels of short-term foreign debt relative to its international reserves might not generate such a run if the country had a strong international liquidity position. In other words, although strong reserve positions can do little to deal with insolvency (fundamental disequilibrium) problems, they are a valuable instrument for countries with merely vulnerable fundamentals.

The effectiveness of high interest rate policies and capital controls has been the subject of considerable

controversy. Because so many factors are typically at work during a crisis it is hard to accurately measure the effects of particular policies. High interest rates may send an ambiguous signal to currency markets, as excessively high interest rates may indicate to the markets that the country is in desperate need of capital — not an encouraging sign if the market is already concerned about the safety of investing in that country (Russia in 1998 provides an example of this dynamic). During the Asian crisis, high interest rates in Indonesia did not keep the rupiah from plummeting, because it was not clear at the time whether the high interest rates reflected tight monetary policy or premiums for inflation and political risk.

The effectiveness of capital controls also depends a great deal on how they are interpreted by the market. Capital controls can be seen as prudential measures taken to protect an otherwise stable economy from overly volatile international capital markets, in which case they can be effective in reassuring foreign investors. Capital controls can also be seen, however, as stopgap measures intended to prevent capital from flowing out of an already risky economy. In this case, the implementation of capital controls can itself be a signal to the market that the economy is in worse shape than previously thought. As with international reserves, the introduction of capital controls seems more promising when the economy is in a vulnerable zone than when it exhibits strong fundamental disequilibriums. The introduction of capital controls by Malaysia in 1998 has been the subject of different evaluations along these lines. Probably the best we can say about the Malaysian case is that the controls proved to be neither the disaster predicted by some critics nor the panacea envisioned by some advocates.

Recent crises also remind us that international financial flows can be fickle and that being the darling of international investors today is not a safe indicator of future stability. Although there remains considerable disagreement among economists about the conditions (if any) under which direct controls on international capital flows can be desirable, there is little disagreement that such flows and financial sectors more generally need to be subject to effective prudential regulation. Such policies need to be

carefully crafted: in many countries, policies have contributed as much or more to the generation of crises (through moral hazard and other perverse incentives) as to good prudential oversight.

One of the most notable characteristics of recent currency crises has been the frequency with which they are associated with efforts to maintain various types of adjustably pegged exchange rate regimes. This phenomenon, which has been labeled "the unstable middle hypothesis," was a basic cause of the breakdown of the regime of adjustably pegged exchange rates of the Bretton Woods system. Few analysts dispute that without effective capital controls such adjustably pegged exchange rate regimes are highly prone to crises in the face of substantial international capital mobility. The debate is whether it is necessary to move all the way to one extreme or the other — hard fixes or relatively free floats — to substantially reduce vulnerability to crisis, or whether moving to crawling bands or managed floats is sufficient.

The role of the IMF in crisis prevention and management has been the subject of considerable controversy. Although far from perfect, its ability to identify emerging crisis situations has been much better than its ability to get countries to adopt the necessary preventative actions; the ability of its "seal of approval" to calm crisis situations has also suffered substantial erosion. The IMF has displayed considerable learning, but efforts at substantial reform of the international financial architecture have been largely unsuccessful, and the IMF's ability to act as a crisis manager and international lender of last resort has, so far, only marginally improved. This has resulted in greater self-help efforts undertaken by many countries, especially in Asia, to accumulate high levels of international reserves at both the country and the regional levels. Crisis liquidity is clearly most efficiently provided at the international level, but the failure of substantial reforms at that level makes these individual and regional efforts quite understandable.

See also asymmetric information; balance of payments; banking crisis; Bretton Woods system; capital controls;

capital flight; capital mobility; contagion; currency board arrangement (CBA); early warning systems; exchange market pressure; financial crisis; foreign exchange intervention; hot money and sudden stops; impossible trinity; International Monetary Fund (IMF); International Monetary Fund conditionality; International Monetary Fund surveillance; international reserves; lender of last resort; speculation; spillovers

FURTHER READING

Angkinand, Apanard, Jie Li, and Thomas D. Willett. 2006. "Measures of Currency Crises: A Survey." Claremont Working Papers 2006 06. Downloadable from http://www.cgu.edu/include/2006 06.pdf. Surveys a number of currency crisis indexes, compares their compositions, and discusses which indexes may be appropriately used under which circumstances.

Blustein, Paul. 2005. *And the Money Kept Rolling In . . . and Out: Wall Street, the IMF, and the Bankrupting of Argentina*. New York: Public Affairs. Intended for a general audience, this book provides an insightful account of the Argentine crisis and the roles played by the IMF and American investors.

Buiter, Willem H., Giancarlo Corsetti, and Paolo Pesetti, eds. 1998. *Financial Markets and European Monetary Cooperation: The Lessons of the 1992 93 Exchange Rate Mechanism Crisis*. Cambridge: Cambridge University Press. Provides a historical perspective on exchange rate stability in Europe and presents an interpretation of the European currency crisis of 1992 93.

Eichengreen, Barry. 1999. *Toward a New International Financial Architecture: A Practical Post Asia Agenda*. Washington, DC: Institute for International Economics. Reviews proposals for reform of the international financial system and national policies to reduce the incidence of currency crises and contagion.

International Monetary Fund, Independent Evaluation Office. 2003. "The IMF and Recent Capital Account Crises: Indonesia, Korea, Brazil." Washington, DC: International Monetary Fund. Downloadable from http://www.imf.org/external/np/ieo/2003/cac/index.htm. Provides a useful critical analysis of the role of the IMF in recent crises and contains numerous references to the relevant literature.

Krugman, Paul. 1996. "Are Currency Crises Self Fulfilling?" *NBER Macroeconomics Annual*. Washington, DC: National Bureau of Economic Research. Critically assesses the tenets and conclusions of the second generation currency crisis models.

———. 1998. "What Happened to Asia?" Electronic document. Downloadable from http://web.mit.edu/krugman/www/DISINTER.html.

MacIntyre, Andrew. 1999. "Political Institutions and the Economic Crisis in Thailand and Indonesia." In *The Politics of the Asian Financial Crisis*, edited by T. J. Pempel. Ithaca, NY: Cornell University Press, 143 162. Offers an interesting comparison between the political attributes of Thailand and Indonesia, and traces their respective responses to the onset of the Asian crisis.

Pilbeam, Keith. 2006. *International Finance*. 3d ed. New York: Palgrave Macmillan. Chapter 17 of this introductory textbook contains an easy to read summary of currency crisis theory and provides some illustrative examples.

Rajan, Ramkishen. 2001. *(Ir)relevance of Currency Crisis Theory to the Devaluation and Collapse of the Thai Baht*. Princeton Studies in International Economics No. 88 (February). Discusses the Thai crisis in the context of various generations of currency crisis models.

Saxena, Sweta C. 2001. "The Changing Nature of Currency Crises." *Journal of Economic Surveys* 18 (3): 321 50. Traces the changes in the currency crisis literature from 1979 through the post Asia crisis models.

Willett, Thomas D. 2007. "Why the Middle Is Unstable." *The World Economy* 30 (5): 709 32. Discusses the political and economic reasons why delayed adjustments tend to make adjustable peg exchange rate regimes particularly crisis prone. It contains numerous references to the theoretical and empirical literature by both economists and political scientists.

Willett, Thomas D., Ekniti Nitithanprapas, Isriya Nitithanprapas, and Sunil Rongala. 2004. "The Asian Crises Reexamined." *Asian Economic Papers* 3: 32 87. Presents an overview of the various theories explaining the 1997 98 Asian crises and reexamines the evidence concerning the causes.

ERIC M. P. CHIU, STEFANIE WALTER, JOSHUA C. WALTON, AND THOMAS D. WILLETT

■ currency substitution and dollarization

Currency substitution refers to a broad set of conditions under which two or more currencies can be used as media of exchange. If currency substitution is a continuum within which domestic and foreign currencies cocirculate, dollarization occurs at one end of that continuum, when the foreign currency   most times, the U.S. dollar   is preeminent in domestic transactions in goods and/or assets. Since 2002, the term *euroization* has also emerged. Dollarization may occur either as a formal decision of governments or as the choice of most economic agents in a country. These are known as de jure (official) and de facto (unofficial) dollarization, respectively.

Currency Substitution Currency substitution may refer to foreign currency, deposits denominated in foreign currency units held in domestic banks, or to deposits of domestic residents at foreign banks. The currency reports of the U.S. Treasury are used to determine where U.S. dollars go abroad, but dollars that travel between countries outside the U.S. are not traced (see Feige et al. 2003.) Moreover, shipments of currency in amounts less than $10,000 are not recorded. Several estimates in a 2003 U.S. Treasury report put $300 billion to $400 billion of U.S. currency outside the country, with 25 percent elsewhere in the Western Hemisphere and about 40 percent in Europe and the former Soviet Union.

Motives for currency substitution can come from seeking money as either a store of value or as a medium of exchange. Often economic agents will hold U.S. currency for store-of-value considerations or for use in larger transactions, while the local currency cocirculates for smaller purchases. This characterizes most of the Latin American experience with currency substitution. At risk of loss of their monies due to confiscatory currency reforms of the past, many households in Eastern Europe in the late 1980s held wealth in foreign currencies because they had no alternative assets available. On the other hand, sometimes foreign currencies are needed to provide media of exchange because local currency is in short supply. This can occur, for example, when the central bank is defending an overvalued exchange rate peg.

Institutional factors play a major role in the evolution of currency substitution. In Latin America, for example, most countries have suppressed or limited the holding of foreign currency deposits for a time, only later to remove restrictions as the demand for foreign currency has grown. This has an obvious impact on the quantity of foreign currency held. Other countries have permitted foreign currency holdings from the start (the most extreme case being Panama's decision to adopt the U.S. dollar as the sole currency). A country's policies with regard to currency substitution can and do change over time. Increasing financial development leads to more alternatives to foreign currency as a store of value and removes the necessity of restrictions on its holding.

Once currency substitution occurs, it can become irreversible through what is sometimes referred to as *dollarization hysteresis*. In short, each agent holds currencies based on what it expects other agents to hold. Once other currencies are introduced, a substantial shock must occur to induce agents to return to holding just local currency (Feige et al. 2003).

Dollarization De jure, or official, dollarization most often occurs after a period of high inflation, when a country either finds itself unable to create a new domestic currency that would hold value or is unwilling to endure the output loss needed to create credibility. Unofficial, or de facto, dollarization could include currency as well as asset substitution. Feige (2003) provides a breakdown of the separate currency and asset dollarization channels and their relative sizes. For example, there is almost no currency substitution in the Former Yugoslav Republic of Macedonia, but foreign-denominated monetary assets in Macedonian commercial banks are 2.5 times the size of domestic-denominated monetary assets (excluding cash). In general, however, currency and asset substitution indexes are highly positively correlated (i.e., currency substitution and de facto dollarization). We therefore focus on de jure, or official, dollarization.

De jure dollarization benefits countries if they are small and well integrated into world markets, and if their capital is highly mobile (Mundell 2003). Many but not all of the smaller countries of Europe that

are not part of the European Monetary Union Andorra, Kosovo, Montenegro, Monaco, San Marino, and Vatican City use the euro; only one (Kosovo) has a population of more than 1 million. Many of the island states of the Pacific use the U.S. dollar; a few use Australian or New Zealand dollars. Other states have granted legal tender status to the U.S. dollar and had their own currency circulate alongside. In Panama, which has the longest such history, the dollar is used for most transactions, but Panamanian coins circulate for minor trade. Panama is also highly integrated into the international capital markets. Guatemala and the Bahamas are other examples of de jure dollarized economies in the Western Hemisphere.

In many instances, de jure dollarization occurs through the creation of a currency board, such as in Argentina, but it need not be so. Ecuador makes an interesting example. The largest country by population to dollarize, it chose to abandon the local currency, the sucre, after that currency had depreciated precipitously at the end of 1999. The government declared the dollar legal tender, and the sucre was removed from circulation. The International Monetary Fund sent a technical assistance mission in support of dollarization.

Although official dollarization has helped many countries by imposing monetary discipline, its history in creating fiscal discipline is more mixed. Domestic currencies can gain revenue for a country through seigniorage; dollarization limits the ability of governments to collect an inflation tax (Willett and Banaian 1995), and so dollarization may remove any temptation for governments to resort to it. As a result, many economists have hypothesized that dollarization should produce fiscal discipline since additional spending would more likely be paid for with taxes. But as long as a government can borrow, fiscal discipline is muted. Moreover, if interest rates are reduced because of dollarization and greater monetary policy credibility, fiscal policy may become *more* lax. Fiscal deficits in dollarized Panama were higher than in nearby Latin American countries between 1970 and 1998 (Goldfajn and Olivares 2000). Additionally, dollarization through a currency board

did not help Argentina's fiscal authorities become more restrained. Therefore another expected benefit of dollarization lower interest rates has not always been realized.

The most ardent supporters of dollarization argue that when it comes from the decisions of individuals and is the result of a competition between note issuers, it provides benefits to an economy. Allowing the choice protects the property rights of households and ensures that their earnings cannot be confiscated by the inflation tax. Panama's adoption of the U.S. dollar as the sole legal tender in 1904 stands as the leading case of dollarization in most analyses.

Although many economists argue that the costs of dollarization are real and the benefits ephemeral, others argue that the benefits can be large only if the commitment to dollarization is made official, either through the abandonment of the domestic currency or through the creation of a currency board arrangement. But perhaps the biggest objection to dollarization is the country's lost pride. Few countries come into existence without some national fervor, and that fervor often seeks expression in historical pictures of the native landscape on paper, in the form of domestic currency. Many new currencies quickly become subject to inflationary pressures, but that seldom stops the next new country from thinking that it can do better and celebrating independence with new fiat money.

See also common currency; currency board arrangement (CBA); discipline; euro; exchange rate regimes; international reserves; money supply; multiple currencies; optimum currency area (OCA) theory; seigniorage

FURTHER READING

Calvo, Guillermo A. 2000. "Testimony on Dollarization." Presented before the Subcommittee on Domestic and Monetary Policy Committee on Banking and Financial Services, Washington, DC, June 22. A critical review of the discussion of using dollars in Latin America and the issues with remitting seigniorage revenues back to those governments.

Feige, Edgar L. 2003. "The Dynamics of Currency Substitution, Asset Substitution, and de Facto Dollarization and Euroization in Transition Countries." *Comparative*

Economic Studies 45 (3): 358 83. Provides a classifica tion of types of dollarization with an application to transition economies.

Feige, Edgar L., Michael Faulend, Velimer Šonje, and Ve dran Šošić. 2003. "Unofficial Dollarization in Latin America: Currency Substitution, Network Externalities, and Irreversibility." In *The Dollarization Debate*, edited by Dominick Salvatore, James W. Dean, and Thomas D. Willett. Oxford: Oxford University Press, 46 70. Contains an overview of the amount of dollarization that has happened in Latin America.

Goldfajn, Ilan, and Gino Olivares. 2000. "Is Adopting Full Dollarization the Solution? A Look at the Evidence." Working paper. Rio de Janeiro, Brazil: Pontifícia Uni versidade Católica do Rio de Janeiro. Reviews the history of complete dollarization in Panama.

Hanke, Steve H. 2003. "Money and the Rule of Law in Ecuador." *Policy Reform* 6 (3): 131 45. Reviews the decision of the Ecuadorian government to give up its local currency in order to solve chronic high infla tion.

Mizen, Paul, and Eric J. Pentecost. 1996. "Currency Sub stitution in Theory and Practice." In *The Macro economics of International Currencies: Theory, Policy, and Evidence*, edited by Paul Mizen and Eric J. Pentecost. Cheltenham, UK: Edward Elgar, 8 43. An overview chapter for a book that studies currency substitution.

Mundell, Robert A. 2003. "Currency Areas, Exchange Rate Systems, and International Monetary Reform." In *The Dollarization Debate*, edited by Dominick Salvatore, James W. Dean, and Thomas D. Willett. Oxford: Ox ford University Press, 17 45. A useful synopsis of how the optimal currency area approach is applied to the question of dollarization.

Schuler, Kurt. 2005. "Some Theory and History of Dol larization." *Cato Journal* 25 (1): 115 25. A number of economists have proposed dollarization in many devel oping economies; this article makes the case for its widespread use.

Willett, Thomas D., and King Banaian. 1995. "Currency Substitution, Seigniorage, and the Choice of Currency Policies." In *The Macroeconomics of International Cur rencies: Theory, Policy, and Evidence*, edited by Paul Mizen and Eric J. Pentecost. Cheltenham, UK: Edward Elgar, 77 95. Uses an optimal inflation tax approach to

determine the effects of currency substitution on sei gniorage revenues.

Williamson, John E. 2003. "Dollarization Does Not Make Sense Everywhere." In *The Dollarization Debate*, edited by Dominick Salvatore, James W. Dean, and Thomas D. Willett. Oxford: Oxford University Press, 172 76. A more skeptical look at the advice to central banks to give up domestic money issues.

KING BANAIAN

■ current account

See balance of payments

■ customs unions

Customs unions are arrangements among countries in which the parties do two things: (1) agree to allow free trade on products within the customs union, and (2) agree to a common external tariff (CET) with respect to imports from the rest of the world. Cus toms unions and preferential trade arrangements more generally have become increasingly important in recent years. The most famous example of a cus toms union is the European Union (EU). Trade among the member states of the EU flows tariff free, and regardless of which country in the EU imports a product, the same tariff is paid. The CET is what distinguishes a customs union from a free trade area. In a free trade area, trade among the member states flows tariff free, but the member states maintain their own distinct external tariff with respect to imports from the rest of the world. The North American Free Trade Agreement is the best known example of a free trade agreement. Canada, the United States, and Mexico do not share a common external tariff, de spite allowing free trade on products traded among the three countries.

With the exception of high protection on agri cultural products and some "sensitive" products, the EU generally has low tariffs, and the competitive markets of the EU in manufactures have been cred ited with improving the economic performance of

the EU countries. The impact of customs unions among developing countries (or "South-South" customs unions) on the development of the participating countries has been ambiguous at best, however. There are many examples of customs unions among developing countries, including Mercosur (Argentina, Brazil, Paraguay, Uruguay, and, as of 2006, Venezuela), the Central American Common Market (Guatemala, Nicaragua, Costa Rica, Honduras, and El Salvador), the Eurasian Economic Community (Russia, Belarus, Kazakhstan, Kyrgyzstan, Tajikistan, and Uzbekistan), the East African customs union (Kenya, Tanzania, and Uganda), and the Southern African Customs Union, which is the oldest customs union in the world, comprising South Africa, Botswana, Lesotho, Namibia, and Swaziland.

Both economic theory and empirical evaluations of customs unions and free trade agreements are ambiguous regarding the usefulness of these arrangements for the growth and welfare of the participants. The reason these agreements can be harmful is that "trade diversion" can occur. Since the agreements are discriminatory, tariffs are paid on imports from the rest of the world but not from partner countries. Private individuals may import a tariff-free product from producers in partner countries, even though the price on world markets (the tariff exclusive price) of the goods is cheaper from suppliers in the rest of the world. For example, suppose a partner country will supply a product at $1.25, while a nonpartner will supply it at $1. If the same tariff is charged on imports from all countries, the home country will import from the cheapest source and pay $1. Under a customs union, however, if the tariff exceeds 25 percent, private individuals will have an incentive to import from partner countries; then the home country will lose the tariff revenue charged on nonpartner imports and will pay more in foreign currency for imports. This is known as trade diversion. When partner countries are the low-cost suppliers in the world, then lowering tariffs preferentially will result in benefits to the importing country from increased imports from the most efficient world suppliers a phenomenon known as "trade creation." Excluding market access considerations, eco-

nomic theory tells us that nondiscriminatory tariff reduction would bring the country the gains of trade creation without the losses of trade diversion, and therefore nondiscriminatory tariff reduction dominates preferential trade arrangements on economic grounds. An additional important result of economic theory is that welfare-worsening customs unions (those where trade diversion dominates) can be made welfare improving if the external tariff of the customs union is lowered sufficiently. For example, in the case just discussed, if the tariff is lowered to less than 25 percent, the home country will import from the most efficient supplier on world markets and avoid the trade diversion.

The first incarnation of the Central American Common Market (CACM 1) is worthy of examination, in part because it resulted in an unusually large expansion of intracustoms union trade from 7.5 percent of total trade in 1960 to about 22 percent between 1976 and 1982 and because it is an example of extensive trade diversion. Using high effective protection on manufacturing goods, the CACM countries significantly expanded their manufacturing industries during this period by selling manufacturing products to one another and reducing their imports of these products from the rest of the world. These products were not competitive on world markets, however, and traditional agricultural exports, such as coffee, remained the primary export products to the rest of the world. That is, the increase in intraunion trade in CACM was accomplished primarily by costly trade diversion. After the debt crisis of the early 1980s, many of these industries and intraunion trade went into decline. CACM 1 used the customs union to focus production inward, that is, for import substitution industrialization, and did not use the customs union as a means of fostering competition.

When the CET of the customs unions results in a significant amount of trade diversion or fails to protect influential domestic industries, countries often exclude these products from the CET of the customs union (such as the agriculture product exclusions in the customs union agreement between the EU and Turkey and the many exclusions in the

Mercosur agreement). In the case of the Eurasian Economic Community, the CET is reportedly applied to less than 60 percent of the tariff lines for most of the countries in the customs union through selective application of the CET without prior agreement on exclusions.

Although Kemp and Wan (1976) showed that it is theoretically possible to design a common external tariff (with lump sum transfers among partner countries) that leaves excluded countries no worse off and at least one partner country better off, numerical and econometric assessments indicate that countries that are excluded from preferential arrangements will typically lose. The reason is that due to the tariff-free intrablock trade, buyers within the preferential trade area divert purchases away from excluded countries toward partner countries. This reduces demand and the price that sellers from excluded countries may obtain in the preferential trade area.

An advantage of customs unions over free trade areas is that customs unions eliminate the need to have "rules of origin" schemes among the member countries. That is, in free trade areas, traders may import a product into a member country with a low tariff and then resell it to another member country with a high tariff. In order to prevent this kind of tariff rate arbitrage, traders must document that the product crossing borders within the free trade area is produced within the free trade area. Rules of origin are then typically established that determine the conditions under which the product may be considered as one produced within the free trade area. These rules of origin can be rather cumbersome and difficult to administer, and they often impede trade.

For a country that desires to use trade policy as a means of expanding exports and competing on world markets, a crucial disadvantage of a customs union relative to a free trade area is that the country cedes the power to lower its tariff to the customs union's tariff-setting authority. Some economic theory and empirical work suggests that tariff-setting authorities in customs unions are less likely to lower tariffs than individual countries. Faced with high-cost imports from a partner country, the individual country in a free trade agreement has the liberty to lower its ex-

ternal tariff to reduce or eliminate this trade diversion. For that reason, Chile refused to join the Mercosur customs union but has vigorously pursued free trade areas with most of its trading partners so that more than 90 percent of its imports enter tariff free, while also lowering its uniform tariff from 11 to 6 percent.

Customs unions are said to provide the advantage of political cooperation, which contributes to peace and security. Peace between France and Germany in the EU is the best example of this. But if there are costs to some of the participants, the customs union could have the opposite impact. The U.S. Civil War was partly motivated by resentment of the southern states over having to pay high prices for northern states' manufacturing products due to high tariffs on European imports. In the East African customs union, the Kenyan manufacturing sector was more developed than that in Tanzania and Uganda. The latter two countries complained of bearing most of the trade diversion costs due to having to pay high prices for manufactured products from Kenya, but they could not independently lower the tariff on these products. Reportedly resentment over the uneven costs of the customs union contributed to border hostilities between Tanzania and Uganda in 1979.

Although theoretically unambiguous conclusions about customs unions are difficult, a key rule of thumb is to use the customs union to increase competition. Crucially, tariff reduction can reduce the costs of trade diversion and foster competition. "Deep integration" that fosters competition is important. Notably, the customs union agreement can be used to reduce nontariff barriers, such as regulatory and administrative barriers to trade, and commitments to increase the rights of foreign investors can provide for competition in key sectors important for economic development, such as business services. Both Schiff and Winters (2003) and Harrison, Rutherford, and Tarr (2003) suggest that another rule of thumb is that developing countries are more likely to gain from customs unions with industrialized country ("Northern") partners. By virtue of their size, Northern countries are more likely to introduce

competition in the markets of developing countries, and Southern countries are more likely to obtain technological advances from trade with Northern countries.

Customs unions form an important feature of the world economy as evidenced in the EU, Mercosur, and other examples of this kind of regional trade agreement. Consequently, the assessment of customs unions remains an important task for both theoretical and applied trade policy analysis.

See also common market; free trade area; multilateralism; regionalism; rules of origin

FURTHER READING

Harrison, Glenn H., Thomas F. Rutherford, and David G. Tarr. 2003. "Trade Policy Options for Chile: The Im portance of Market Access." *World Bank Economic Re view* 16 (1): 49 79. A numerical assessment, through a global computable general equilibrium model, of the impact of various regional trading arrangements in volving Chile on partners and excluded countries.

Kemp, Murray, and H. Wan. 1976. "An Elementary Prop osition Regarding the Formation of Customs Unions." *Journal of International Economics* 6: 95 97. A formal treatment of conditions under which a customs union can enhance efficiency.

Schiff, Maurice, and L. Alan Winters. 2003. *Regional In tegration and Development*. Washington DC: World Bank and Oxford University Press. An excellent, clearly written, comprehensive treatment of theory and practice of preferential trading areas.

DAVID G. TARR

D

■ debt deflation

Debt deflation can occur if falling prices raise the real costs of repaying loans, thereby boosting the costs of debt service and leading to higher bankruptcy rates and debt defaults. Rising debt defaults then produce increasingly weaker business conditions, further fueling the downward spiral and leading to a vicious circle that the economist Irving Fisher believed to be an important element prolonging and deepening the Great Depression of the 1930s in the United States. Fisher's perspective represented a sharp departure from the earlier conventional wisdom that occasional deflation was a natural result of productivity gains and, as such, was to be more welcomed than feared. Such was clearly not the case with the contracting economy of the early 1930s, for which Fisher's mechanism emphasized the link between declines in goods prices and declines in asset prices. Although economists often discount the role played by the October 1929 Wall Street crash, under Fisher's debt-deflation mechanism an event like this serves as a catalyst for defaults that could, if widespread enough, induce deflation in the economy as a whole.

The role of balance sheet effects is also important in fueling a debt-deflation process (Bernanke 1983). Debtors who default forfeit their assets to banks. Sudden, large drops in the prices of the forfeited assets then hurt bank balance sheets and potentially threaten the solvency of the banking sector. If banks curtail lending as a result, firms dependent on bank credit could face a credit crunch that leads to further production cutbacks and intensifies the deflationary spiral. Firms' borrowing difficulties may be further exacerbated if declines in asset prices reduce the value of their loan collateral.

Country Experiences with Deflation During Japan's slide into deflation after 1989, land and equity price declines hurt firms' loan collateral at the same time that the banks' own direct exposure to the stock market weakened their balance sheets. Together, these factors made firms less creditworthy at the same time that banks became more reluctant to lend at all. Japan in 1989, like the United States in 1929, clearly suffered from a sudden decline in asset prices as the stock market crashed. In the U.S. case, the unprecedented debt buildup during the 1920s, coupled with a sudden shift from a stable price environment to a deflationary environment at the beginning of the 1930s, meant that firms and households may have faced an unexpected rise in their debt service costs at the very time that their ability to fund their debt diminished. The unanticipated nature of the rising real debt burden at the beginning of the 1930s would have been a key factor in the operation of a debt-deflation process during the Great Depression in the United States (Fackler and Parker 2005). The agricultural sector was especially vulnerable in the U.S. case, and as falling prices made it harder and harder for farmers to repay loans, banks dependent on farm credit suffered as well. It is much harder to make a case for the importance of the debt-deflation process outside the United States, however, given that other major economies such as the United Kingdom did not appear to face any such dramatic debt buildup during the 1920s.

In addition to the more familiar U.S. and Japanese experiences, a debt-deflation mechanism may have been at work in late 20th-century emerging market crises such as those experienced in Mexico, Russia, and Southeast Asia. Output declines can be fueled by credit constraints that limit firms' access to working capital. These binding credit constraints in turn trigger debt deflation under this approach, leading to the dumping of assets and falling asset prices that further tighten the existing credit constraints. Once under way, deflation then lowers the marginal product of and real rates of return on factors of production. Vulnerability to this process increases with leverage, and economist Enrique Mendoza (2006) points to a surge in leverage ratios (as reflected in the ratios of debt to firm sales, book value of firm equity, and market value of firm equity) of listed corporations in Indonesia, Korea, Malaysia, and Thailand in the period leading up to the 1997 Asian financial crisis. As with the earlier debt-deflation literature, the unfolding of this process requires a high debt buildup. Otherwise the binding collateral constraint would not come into play.

While the actual prevalence of debt deflation remains a subject of debate among economists, almost all would consider the potential triggers, or facilitating factors, for such a process undesirable. That is, it is hard to see how sudden asset price declines, the sudden onset of deflation, or high debt levels and leverage ratios could be considered desirable policy goals. The Federal Reserve in the 1920s and the Bank of Japan in the 1980s, in arguably fostering an extended period of easy credit policies followed by sudden tightening, may well have played a major role in producing these very conditions. The pattern of sudden collapse following an extended period of excess liquidity seems to have been repeated in Southeast Asia in the 1990s. Unfortunately, such patterns often become more obvious after the fact than they were before the declines began.

Moreover, even though sudden tightening may trigger a debt-deflation process, expansionary monetary and fiscal policy will not necessarily reverse it in the absence of expectations of recovery on the part of consumers, businesses, and banks. Consumer and business unwillingness to spend, and bank unwillingness to lend, may well become ingrained among survivors of a debt-deflation process. In the Japanese case, for example, the bad debt problems of many Japanese corporations and the risks of future loan defaults help to explain why, even with ample zero-interest-rate money available, banks remained reluctant to lend. At the same time, Japanese banks were seeking to bolster their own balance sheets, which had been hurt not only by nonperforming loans but also by the sharp drop in the market value of their equity holdings. The banks' reduced willingness to lend and to circulate the new money being created by the Bank of Japan meant that these funds were often simply held within the banking system, thereby doing little to fuel new spending and combat ongoing deflation in the second half of the 1990s.

Prevention May Be Easier Than Cure Debt-deflation concerns only add to the argument that the most effective way to fight deflation is to keep it from starting, or at least to keep it from lasting long enough to become entrenched in expectations. Indeed, once consumers come to anticipate continued falls in prices, they have an incentive to postpone nonessential purchases, to save now and consume later. Anticipation of deflation also encourages people to accumulate cash balances in the hope that their purchasing power will grow over time. All this further reduces the current demand for goods and services, adding to the problems that businesses face. And, once deflation has set in, even a zero *nominal* interest rate policy, as the Bank of Japan maintained for some years, cannot prevent real borrowing costs from rising in the face of further price declines. This reflects the fact that the purchasing power of the funds borrowed increases over the course of the loan, producing a gain for the lender and a loss for the borrower even if no nominal interest payment is collected. Indeed, the worse the deflation gets, the more the automatic rise in the *real* interest rate is likely to further curtail business spending and exacerbate the downward pressures on economic activity.

There certainly are a number of reasons for authorities in low-inflation environments to balance the risks of deflation against those of accelerating

inflation. The onset of declining prices can be particularly dangerous for a highly indebted economy vulnerable to debt deflation. Sudden interruption of a long period of easy credit and rising asset prices certainly had dire consequences for the United States at the end of the 1920s and Japan at the end of the 1980s. The somewhat analogous boom-bust cycle in the emerging market economies of Southeast Asia in the 1990s, although it did not produce sustained deflation of goods prices, nevertheless led to an extended period of decline in both output and asset prices following the outbreak of the Asian financial crisis in 1997.

See also balance sheet approach/effects; currency crisis; Federal Reserve Board; financial crisis; hot money and sudden stops; liquidity trap, the; money supply; seigniorage

FURTHER READING

Bernanke, Ben S. 1983. "Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression." *American Economic Review* 73 (3): 257 76. Points to the negative effects of the 1930 33 U.S. banking crisis on credit allocation during the Great Depression.

Burdekin, Richard C. K., and Pierre L. Siklos, eds. 2004. *Deflation: Current and Historical Perspectives.* New York: Cambridge University Press. Includes an overview of deflationary processes in chapter 1 and detailed analysis of the Japanese case by Michael Hutchison in chapter 9.

Fackler, James S., and Randall E. Parker. 2005. "Was Debt Deflation Operative during the Great Depression?" *Economic Inquiry* 43 (1): 67 78. Argues that a large, unanticipated increase in the real debt burden in the United States at the beginning of the 1930s provided the catalyst for debt deflation.

Fisher, Irving. 1933. "The Debt Deflation Theory of Great Depressions." *Econometrica* 1 (4): 337 57. Fisher's original exposition of a debt deflation process in which over indebtedness is followed by attempted liquidation that leads to falling prices and still higher real debt burdens.

Mendoza, Enrique G. 2006. "Lessons from the Debt Deflation Theory of Sudden Stops." *American Economic Review* 96 (2): 411 16. Points to debt deflation being operative in recent emerging market crises and triggered by credit constraints imposed on highly indebted firms.

RICHARD C. K. BURDEKIN

■ democracy and development

Although there is a strong positive association across countries between democratic governance and economic development, the exact nature of the relationship between these two concepts is less than perfectly clear and has been the subject of extensive study and intense debate. This entry will describe mechanisms by which democracy may or may not promote development, mechanisms by which development may or may not promote democracy, and causal effects from other variables on both development and democracy.

Democracy has many definitions, but recent scholarly literature has concentrated on variants of the following: democracy is the form of government in which political leaders are selected by periodic elections conducted reasonably honestly under conditions in which all political interests can compete. In addition to the right to vote, citizens have the rights to form political organizations, to try to influence public opinion, to put pressure on officeholders, and to compete for offices. This competition must take place under conditions of free expression through the media and the right of peaceful assembly. Note that this definition does not imply that all groups in the population, in particular the poor, are effectively represented in governmental decision making. Whether democracy results in implementation of the preferences and interests of the majority of citizens is an empirical question, not part of this definition of democracy.

Two widely used empirical measures of democracy are those of Freedom House and Polity IV. Since 1972, Freedom House has provided annual ratings of political rights and civil liberties for most of the countries of the world. Polity IV has combined measures of the competitiveness and openness of executive recruitment, the competitiveness of political

participation, and constraints on the executive into annual scores of democracy and autocracy for some 140 countries; Polity IV's data go back to 1800 for some countries. Przeworski et al. (2000) constructed a dichotomous annual measure of democracy for the period 1950 90; their definition adds the requirement that an incumbent party has lost an election and turned over power to the opposition. Despite the differences in definition, there are generally high correlations among these and other similarly inspired empirical measures of democracy.

Democracy Promotes Development A considerable literature in institutional economics contends that the economic rise of the West is inextricably linked to the economic institutions of secure property rights and third-party contract enforcement through impartial courts, and that these economic institutions were founded on political institutions of separation of powers and representative government (North 1995). The contention is that autocratic governments, unconstrained by constitutional limitations backed by independent legislatures and courts, failed to provide secure property rights over the long term. The political institutions under which modern economic growth emerged in Europe and its offshoots (especially North America and Australasia) allowed the market system to flourish and the business class to invest and innovate, fairly secure from government predation and social unrest. The regimes also began to supply free public education, which further contributed to economic progress (Lindert 2004). The fact that these political regimes were able to provide a good investment climate despite substantial restrictions on the franchise owes something to the legacy of political and social deference by the lower classes, a deference that gradually disappeared. Prior to World War I the partial democracies of Europe and its offshoots were able to accommodate the rising demands of the middle and lower classes without a great deal of social unrest or harsh repression. By 1950 the political attitudes that tolerated restrictions on the franchise had virtually disappeared; in the postwar period, democracy essentially implies full adult suffrage.

This institutional economics literature claims that democracy promotes economic development by providing secure property rights, impartial third-party contract enforcement, social peace, and public education. There are exceptions to these generalizations, however, as some autocratic regimes have expanded educational opportunities far faster than the typical democracy, and a great deal of economic expansion has taken place under nondemocratic rule, presumably with secure enough property rights to encourage investment and innovation. A few autocracies in East Asia, namely South Korea, Taiwan, and Singapore, created institutions and followed policies that have fostered economic miracles, raising their populations from poverty to middle-income status in four decades, and it is arguable that their autonomy from democratic pressures was instrumental in their rapid growth in the early stages of industrialization. All of these regimes expanded educational opportunities and physical infrastructure more rapidly than similarly situated democracies. Strong bureaucracies provided contingent rents to entrepreneurial activities and withdrew privileges if performance did not match targets. Most other autocracies in the decades since World War II have not provided an environment conducive to growth, or they did so only for brief periods.

Well-functioning democracies have links between civil society and political parties that translate popular desires into government programs (Lipset and Lakin 2004) and civil services that can implement these programs. Countries meeting the definitional requirements of democracy may lack these institutional features of good government and fail to provide physical infrastructure, secure property rights, high-quality public education, and macroeconomic stability.

Economists have examined how inequality and type of regime affect economic growth. A high level of income and wealth inequality may promote growth by putting income into the hands of people who will save and take entrepreneurial risks, but inequality probably also inhibits growth through its effects on sociopolitical instability and inequality's adverse effects on the level of human capital (because

of the inability of credit-constrained families to finance highly productive investments in human capital and because of the unwillingness of the rich to finance public education). Many models (e.g., Benabou 2000) incorporate the idea that democracy reduces inequality through redistributive fiscal policy; this redistributive policy has ambiguous effects on growth in the models. It inhibits growth through distortion of labor supply, saving, and risk-taking, but it promotes growth through its effects on human capital accumulation, especially if human capital generates positive externalities. The empirical literature finds little support for the proposition that democratic redistribution impairs growth through distortion of effort and saving incentives (Lindert 2004), but human capital accumulation probably does promote growth. A remaining question is whether the models are correct in assuming that democracy does in fact bring about redistribution of the growth-promoting variety. We take up this question later.

The econometric literature on economic growth in the postwar period has yielded mixed results on the effects of democracy on economic growth. No study prior to 1987 found in favor of democracy, and no study after 1988 found in favor of dictatorship (Przeworski et al. 2000, 178), but it is clear that growth can occur under either type of regime. Przeworski et al. (2000) find that democracy causes somewhat more rapid growth in output per worker, but regime type makes no difference to growth in very poor countries. There is some evidence that democracy reduces income inequality (Tavares and Wacziarg 2001), which is probably conducive both to economic growth and to the survival of democracy, but the same study finds that democracy inhibits investment, with a pronounced negative effect on growth.

The ambiguous effects of democracy on growth seem to be the result of multiple causal chains that offset one another. One of the ways democracy affects growth is through its effects on political instability. A regime change, such as from democracy to autocracy or vice versa, or the replacement of one dictator by another through unconstitutional means,

seems to have an adverse effect on growth, at least in the short run. At low levels of income regime change is no less likely under democracy than under autocracy, but as income level increases (either over time or in the cross-section) the failure of democracy becomes progressively less likely. Another aspect of political instability is social unrest, as manifested in strikes, antigovernment demonstrations, and riots. These events are actually more common under democracy than under dictatorship, but their occurrence has much more adverse effects on growth under dictatorship than under democracy (Przeworski et al. 2000). In other words, democracy accommodates this kind of social unrest without serious adverse effects on growth, but dictatorships in general do not.

Another channel through which democracy may affect economic growth is education. Many authors have asserted that extension of the franchise to all the people results in governmental support of mass education. This seems a very plausible statement when one looks at the spread of primary education and the franchise in Europe and the New World in the years before World War II (Lindert 2004). A landed elite, especially one that is ethnically different from the bulk of the population (as was the case in much of Latin America and the Caribbean), would not want to tax itself to provide mass education and might well fear that such education would threaten its hold on power. In fact, the extension of the franchise and the growth of primary education are fairly highly correlated across these countries in this period. The argument is less clear as applied to the postwar world, for while there is some econometric evidence showing a positive effect of democracy on primary and secondary enrollment rates while controlling for income level (e.g., Tavares and Wacziarg 2001), the evidence of a large effect is not strong. Moreover, one careful study failed to find a significant effect of democracy on the share of public expenditure on education in national output (Mulligan et al. 2004). This finding is not surprising, given the way democracy is defined in these studies.

Some countries that meet the empirical definition of democracy are so lacking in governmental capacity and in political parties and electoral institutions that

reflect the popular will that the public investments in human capital and infrastructure are not forthcoming. But if one adopts a more substantive concept of democracy (as in, for example, Tilly 2007), it seems safe to say that fully democratic regimes will provide more of these investments than the average autocracy (leaving aside that small number of autocratic "developmental states" in East Asia, some of which have now become democracies).

In a surprising finding, Przeworski et al. (2000) note that fertility is lower under democracies than under dictatorships, after adjusting for many factors that affect fertility. One of these factors is the parents' desire for security in their old age, which might be provided by having a large number of children, or by having a smaller number of children who are more likely to survive and earn enough to assist their parents, or by accumulating financial assets or pension rights. Przeworski et al.'s interpretation is that democracy provides greater confidence in the continuity of economic policies, as well as lower infant and child mortality, and thus tips the balance in favor of having fewer, better-educated children rather than many offspring.

Development Promotes Democracy A large literature in political science and political sociology attributes the emergence and spread of democracy to various aspects of economic development. In traditional society a landholding or bureaucratically based elite is able to hold the reins of political power and preserve its privileges, but as development occurs, structural changes weaken the power of the traditional elite. Wealth accumulates in the hands of industrial and commercial interests, rural people move to the cities, an industrial labor force grows in factories and mines, and education expands at both primary and higher levels. Some scholars (e.g., Lipset 1959) contend that, in addition to these structural changes in the economy, modernization associated with economic development increases the level of rationality in individual decision making, encourages toleration of opposing points of view, and attenuates the struggle over economic resources. A pervasive theme in much of the literature is that, whether through structural changes or changes in political

culture, economic development makes authoritarian rule less viable and promotes the emergence and survival of democratic regimes. Strong versions of this perspective would contend that democracy is unlikely to survive in very poor countries and becomes inevitable as countries become highly developed. This modernization theory was heavily criticized as many relatively well-off democracies (especially in Latin America) succumbed to military coups and authoritarian rule and many other countries grew rich without showing any signs of democratization. The theory enjoyed a resurgence in the 1990s as the "third wave" of democratization around the world became manifest, and the empirical studies of that decade confirmed the association of income and democracy (Lipset and Lakin 2004).

An alternative view is that, rather than development leading more or less automatically to democratization, countries are on different paths of institutional and economic development. As a result of historical factors, some countries developed political institutions of constitutional government that encouraged economic development and led ultimately to full democracy, while other countries got onto paths of continuous or intermittent autocracy that have not evolved into stable democracy, despite considerable levels of economic development. A prominent idea is that experience with electoral contestation under a limited franchise established an institutional basis of political parties and civil society organizations that supported well-functioning democracy. As mentioned earlier, the representative governments of Europe and its overseas offshoots followed such a path, although less successfully in Latin America than in the United States, Canada, Australia, and New Zealand. Among the former colonies that achieved independence after World War II, there is a remarkable association between British colonial status and uninterrupted democracy, which is probably attributable to these colonies' experience with electoral contestation and independent judiciaries under British tutelage. Around 2000, there were eleven low-income countries   that is, with incomes less than Costa Rica's   that had been

continuously democratic since independence; all were former British colonies.

A complementary approach that refines elements of modernization theory links democracy to structural changes in the economy that affect the power relationships among class-based interests. Democracy becomes more probable as the land-owning aristocracy loses political power to industrial interests and the middle class. An important study by Rueschemeyer, Stephens, and Stephens (1992) contends that in Europe, Latin America, and the Caribbean the working class was decisive in the push for full democracy. Some scholars doubt that interpretations based solely on the conflict of social classes are adequate explanations of democratic transitions; they stress elite choices and elite consensus on democratic rules, relatively independent of class interests. Collier (1999) analyzes the role of working-class organizations in Europe and Latin America and presents a rather mixed picture, with elites playing the dominant role in some of the cases.

In contrast to the nonformal approach of most of the sociological and political science literature, some economists and analytically minded political scientists have developed formal models of the determinants of democracy. A central question in these models is, under what circumstances will an authoritarian ruling social class decide to extend political rights to other social classes? In one type of model (Bourguignon and Verdier 2000), political participation depends on the individual's level of education; thus a decision to finance education for the lower classes may threaten the elite's political control. The elite benefits economically from the external effects of educating the lower classes, however, and hence may decide to finance public education, despite the fact that the new, larger electorate may impose redistributive fiscal policy to the detriment of the elite. In this model, the greater the initial inequality of income, the less likely is the elite to pursue a policy of public education that leads to a more democratic regime. In another type of model (Acemoglu and Robinson 2006), the ruling elite faces occasional threats of disruption or revolution by the middle or lower classes, who are demanding some redistribu-

tion of society's resources. These classes would not be satisfied with a promise by the elite to redistribute, because they know that the threat of disruption or revolution is temporary, and the elite could renege on their promise once the emergency has passed. The installation of democratic institutions, however, permits the elite to commit to redistribution and thereby avoid both the costs of repression and the danger of revolution. This framework gives rise to a very rich set of models describing precise mechanisms by which economic structures and social stratification may or may not lead to democratization and the survival of democracy.

The Econometric Literature By the 1990s abundant data had become available on national accounts and other economic and social indexes and on measures of democracy covering various periods from the 1950s to the present. Barro (1999) used democracy measured at five-year intervals as the dependent variable; independent variables included economic and social characteristics such as income and education and also the lagged value of democracy. His regression results suggest that a country's level of democracy will converge gradually to a level determined by the economic and social variables. The positive coefficients on income and life expectancy (viewed as a measure of standard of living) are consistent with the idea that these variables cause democracy to be more likely. In this particular study education does not have a statistically significant coefficient (but of course it is highly correlated with the other variables). A dummy variable for countries heavily dependent on oil exports has a negative sign, indicating that high income produced by petroleum exports does not have the same positive effect on democracy as income earned from other sources. It makes sense that the struggle to control large amounts of unearned government income would imperil democratic government.

Przeworski et al. (2000) studied the frequency of transitions from democracy to authoritarian rule and vice versa, and they related these frequencies to various country characteristics. A notable finding is that the probability of transition from authoritarian rule to democracy is only modestly positively related to

income, but the transition from democracy to authoritarianism declines sharply as income rises. In fact, above an income level of Argentina in 1975, there were no transitions away from democracy during that period (or since). The authors conclude that the positive observed correlation between income and democracy at a point in time is primarily due to the fact that once democracies attain a moderately high level of income, they seldom fail, rather than to a tendency for countries to transition to democracy at higher rates as they become richer. They note that countries can remain under authoritarian rule for many years after attaining middle- or high-income status. (Their statement would have been even stronger if they had included six high-income oil-producing authoritarian countries in the Middle East.) These authors also find an independent effect of education: at each income level, democracies with higher levels of education are less likely to fail. They also find that a high level of income inequality or a small labor share in value added in manufacturing increases the failure rate of democracies.

Acemoglu et al. (2005) have presented an empirical challenge to the received wisdom that income causes democracy. The challenge is based on an econometric technique in which the estimated effect of income on democracy is determined solely by variations over time within countries. They find that changes in a country's income do not predict changes in that country's level of democracy. Their interpretation of the data is that prior to the Industrial Revolution some countries developed political institutions that constrained their monarchs and limited the powers of the elite, and these countries evolved along a path of prosperity and increasing democracy, while other countries with different historical characteristics evolved under autocratic and elite-dominant institutions. They report that for various periods from 1840 to the present, there is no tendency for countries that grew more rapidly to experience greater improvements in their democracy scores, compared to countries that grew more slowly. Similarly there is no tendency for countries with greater increases in education to experience greater improvements in their democracy scores. In those parts of the world where the disease environment encouraged European settlement and the displacement or extinction of the local population, the settlers demanded and obtained representative institutions that evolved into democracy, but in the tropics, where death rates of Europeans were high, the Europeans set up institutions designed to extract resources from the colonies for the benefit of the mother country. These extractive institutions were conducive neither to economic growth nor to an evolution toward democracy. They support their interpretation by noting that after adjusting for the rates of settler mortality in the 19th century and population density in 1500, they find that the apparently strong positive effect of income on democracy is severely attenuated.

The interactions between democracy and development have been the subject of much theoretical and empirical investigation. The recent availability of comprehensive data sets has permitted careful statistical studies, which have challenged interpretations derived from casual empiricism. The processes are complex, data limitations persist, and much remains unresolved, however. Scholarly opinion on causal effects is likely to continue to evolve, not least because the world will continue to throw up unexpected developments.

See also corruption; development; economic development; international institutional transfer; political economy of policy reform

FURTHER READING

Acemoglu, Daron, Simon Johnson, James A. Robinson, and Pierre Yared. 2005. "Income and Democracy." NBER Working Paper No. 11205. Cambridge, MA: National Bureau of Economic Research. Presents evidence against the popular hypothesis that economic growth leads to democratization.

Acemoglu, Daron, and James A. Robinson. 2006. *Economic Origins of Dictatorship and Democracy*. New York: Cambridge University Press. Presents a game theory model that spells out causal determinants of democ

racy. Contains case studies and an excellent review of literature.

Barro, Robert J. 1999. "The Determinants of Democracy." *Journal of Political Economy* 107: S158 S183. An important econometric study linking democracy to income, education, and a history of democracy.

Benabou, Roland. 2000. "Unequal Societies, Income Distribution, and the Social Contract." *American Economic Review* 90 (1): 96 129. Theoretical model of the evolution of inequality and democracy.

Bourguignon, François, and Thierry Verdier. 2000. "Oligarchy, Democracy, Inequality, and Growth." *Journal of Development Economics* 62: 285 313. A theoretical model in which the elite may be willing to trade political power for greater economic gains.

Collier, Ruth Berins. 1999. *Paths toward Democracy: The Working Class and Elites in Western Europe and South America.* New York: Cambridge University Press. Study by a political scientist of the spread of the franchise and democracy.

Lindert, Peter H. 2004. *Growing Public: Social Spending and Economic Growth since the Eighteenth Century.* An economic historian looks at the spread of education and democracy. Considers a less procedural and more substantive definition of democracy.

Lipset, Seymour Martin. 1959. "Some Social Prerequisites for Democracy: Economic Development and Political Legitimacy." *American Political Science Review* 53: 69 105. Classic article presenting a version of modernization theory.

Lipset, Seymour Martin, and Jason M. Lakin. 2004. *The Democratic Century.* Norman, OK: University of Oklahoma Press. Mature thoughts of an eminent political sociologist on political organization, civil society, political culture, and democracy in Latin America and the United States.

Mulligan, Casey B., Richard Gil, and Xavier Sala i Martin. 2004. "Do Democracies Have Different Public Policies than Nondemocracies?" *Journal of Economic Perspectives* 18 (1): 51 74. An empirical study that finds that democracy does not have much effect on economic and social policies.

North, Douglass C. 1995. "The Paradox of the West." In *The Origins of Modern Freedom in the West*, edited by R. W. Davis. Stanford, CA: Stanford University Press, 7 34. Argues that modern economic growth emerged in the West because of constitutional limitations on government power.

Przeworski, Adam, Michael E. Alvarez, Jose Antonio Cheibub, and Fernando Limongi. 2000. *Democracy and Development: Political Institutions and Well Being in the World, 1950 1990.* New York: Cambridge University Press. Careful statistical study of determinants and consequences of democracy. Overturns much conventional wisdom.

Rueschemeyer, Dietrich, Evelyn H. Stevens, and John D. Stevens. 1992. *Capitalist Development and Democracy.* Chicago: University of Chicago Press. This study by a sociologist and two political scientists argues that the working class, more than the middle classes, has played an important role in pushing for democracy.

Tavares, Jose, and Romain Wacziarg. 2001. "How Democracy Affects Growth." *European Economic Review* 45: 1341 78. An econometric study of the channels through which democracy affects development.

Tilly, Charles. 2007. *Democracy.* New York: Cambridge University Press. A distinguished sociologist describes social processes that move polities toward or away from genuine democracy.

CHRISTOPHER CLAGUE

■ dependency theory

Dependency theory argues that the process of underdevelopment in the global economic periphery is intrinsically connected to the process of development in the center. Dependency theory appeared in the 1950s as a critical reaction to the conventional approaches to economic development that emerged in the aftermath of World War II. There are two dependency theory traditions (Dos Santos 2002). The first is the Marxist tradition, influenced by Paul Baran and Paul Sweezy and developed by André Gunder Frank with important ramifications in the works of Samir Amin, Theotônio dos Santos, Arghiri Emmanuel, and Aníbal Quijano. The second dependency tradition is associated with the structuralist

school that builds on the work of Raúl Prebisch, Celso Furtado, and Aníbal Pinto at the Economic Commission for Latin America and the Caribbean. This structuralist approach is best represented by Fernando Henrique Cardoso and Enzo Faletto and by the subsequent contributions from Peter Evans, Osvaldo Sunkel, and Maria da Conceição Tavares. Other schools of thought were heavily influenced by dependency theory and espouse, in some respects, very similar views   in particular, the so-called world-systems tradition of Immanuel Wallerstein and his followers (Topik 1998).

Both groups would agree that at the core of the dependency relationship between center and periphery is the inability of the periphery to develop an autonomous and dynamic process of technological innovation. The lack of technological dynamism and the difficulties associated with the transfer of technological knowledge lead to underdevelopment of the periphery with respect to the center. The main contention between the two groups was ultimately related to the possibilities of economic development in the periphery. Marxists would argue that development in the periphery   meaning fundamentally catching up with the center   was impossible, while structuralists would argue that dependent development was feasible.

Vigorous growth in some parts of the developing world in the 1950s and 1960s seemed to justify the views of the latter group. The enduring process of stagnation after the 1980s debt crisis, however, has led to a reconsideration of the relevance of dependency situations. In particular, some authors argue that a new form of dependency has emerged, one in which technological backwardness and the international division of labor are of secondary importance, and the real obstacle to development is financial dependency, reflected in the inability of peripheral countries to borrow in international markets in their own currencies (Vernengo 2006). The following section discusses the main differences and similarities between the two dependency traditions, and the last one analyzes the financial dependency literature.

External versus Internal Limits to Development For Baran and other Marxists the origins of the center-periphery relation were strictly technological and determined by the international division of labor. The center produced manufactured goods for itself and the periphery, while the latter produced commodities mainly for the center, as well as maintaining a relatively large subsistence sector. Marxist *dependencistas* explained the lack of dynamism in the underdeveloped world as the result of its particular insertion in the world economy. In this view, the process of development depended on capital accumulation, which, in turn, hinged on surplus extraction. A larger surplus led to more accumulation of capital and a higher growth rate. Furthermore, for Marxists it was in the uses of the surplus that the differences between developed and underdeveloped regions were most evident. In the most backward countries, where the process of industrialization did not take hold and agriculture was still dominant, underdevelopment resulted from the patterns of land tenure that led to excessive concentration of ownership.

The predominance of large estates in plantation societies implied that a great part of the surplus remained in the hands of landowners, who emulated the consumption patterns of developed countries. Excessive and superfluous consumption of luxuries then reduced the potential for investment and capital accumulation. Conspicuous consumption implied that the surplus was not reinvested in the periphery, but ended financing investment in the center. Hence, conspicuous consumption was the cause of stagnation in the periphery. The international division of labor, which promoted the export-oriented plantation system in a large part of the developing world and reinforced the need for luxury imports, was at the core of the dependency relation.

If industrial development took place, then a new pattern of dependency would emerge. Industrialization would take place with participation of foreign capital, which would tend to control domestic markets. The periphery then would jump into the monopolistic phase of capitalistic development. The surplus extracted by monopolistic capital would not be reinvested in productive activities in the host country, however. Part of it would simply be sent abroad as profit remittances, while the other

part would be spent on conspicuous consumption. Frank (1967) concluded that the only way to break with the circle of dependency would be a political revolution.

The significant economic development in a good part of the periphery during the 1950s and 1960s led to a critique of the Marxist tradition by structuralist authors. Cardoso and Faletto (1967) argued that not only was capitalist development in the periphery possible, but foreign capital also had a tendency to be reinvested in the host country so that foreign investment might in fact stimulate domestic investment. Hence the nature of dependency was such that partial or dependent development was viable. As a result, dependency was not a relationship between commodity exporters and industrialized countries, but one between countries with different degrees of industrialization.

Cardoso and Faletto also distinguished between political and economic variables in explaining dependent development. Development and underdevelopment were economic categories related to the degree of development of the productive structure, and to its level of technological advancement. On the other hand, dependency and autonomy referred to the degree of development of the political structure, and the ability of local political elites to take economic decision making into their own hands. As a result, dependent development in association with foreign capital was possible and occurred in countries such as Argentina, Brazil, and Mexico, and in parts of East Asia, one might add. These were the countries that corresponded to what world-systems authors refer to as the semiperiphery. These were countries in which the simple dualistic vision of a center and periphery, where development was impossible in the latter group, did not seem to fit the real experience.

Cardoso and Faletto emphasized the importance of domestic internal developments, in contrast to the external forces of the world economy, as the main determinant of the situation of dependency. It was the internal political process that led to outcomes that favored foreign actors in the process of development. Furthermore, national capitalist development was not incompatible with the absorption of technological knowledge from multinational firms. Arguably, if the goal was to achieve development, dependent development was a reasonable road to it, even if autonomous development was politically more interesting.

In refuting the Marxist emphasis on the relevance of external factors, however, the structuralist version of dependency went to the other extreme and claimed that internal forces were the almost exclusive determinant of development. The inability to generate technical progress domestically, the domestic patterns of consumption, and the limitations of the domestic elites that opted for political dependency were to blame. If the successful industrialization of some parts of the periphery showed the weakness of the Marxist tradition, then the debt crisis of the 1980s and the failure to renew the process of development in the 1990s proved that the optimism of the structuralist approach was not necessarily warranted.

Financial Dependency and the Original Sin
The debt crises of the early 1980s indicated that the development possibilities in the periphery were not limited just by the inability to catch up technologically. Tavares (2000) argues that the technological division of labor in which the periphery concentrates in the production of commodities for the center, while the latter produces manufacturing goods for the former is of very limited historical relevance. Industrialization and technical progress in the periphery were not sufficient to break the dependency ties with the center. Financial dependency is reflected in the inability of peripheral countries to borrow in international markets in their own currencies and constitutes the real obstacle to development. The new interpretation of dependency situations puts "international money and not technical progress as the expression of financial capital domination over the periphery in the last 150 years" (Tavares 2000, 131 32).

The inability of domestic capitalists and governments to borrow in international markets in their own currencies reflects the inability of the domestic currencies of peripheral countries to acquire all the functions of money, as reserve of value, unit of account, and medium of exchange. The ability to

function as international money is a question of degree. Cohen (1998) suggests that there is a pyramid that reflects the geography of money, with internationalized currencies at the top and fragile currencies, on the verge of currency substitution, at the bottom. The main problem associated with the inability to provide all the monetary functions is that financial markets remain underdeveloped in peripheral countries, and the process of capitalist accumulation is hindered.

Mainstream economists have also dealt with financial dependency. Eichengreen, Hausmann, and Panizza (2003), following previous contributions by Hausmann, argue that underdevelopment results in part from the so-called original sin, that is, the fact that the currencies of developing countries are inconvertible in international markets. In this view, the external instability of domestic currencies in the periphery hinders the process of capital accumulation. Although mainstream and dependency authors agree on the importance of currency inconvertibility, they disagree on the solutions. Mainstream authors emphasize the importance of sound fiscal policies and monetary rules that promote credibility, while dependency authors emphasize the need for capital controls and reduced integration with international financial markets.

See also original sin; Washington consensus

FURTHER READING

Cardoso, Fernando Henrique, and Enzo Faletto. 1967. *Dependency and Development in Latin America*. Berkeley: University of California Press, 1977. An interdisciplin ary book that argues that associate dependent develop ment results from political rather than economic factors.

Cohen, Benjamin. 1998. *The Geography of Money*. Ithaca: Cornell University Press. An analysis of the power structure in international currency markets.

Dos Santos, Theotônio. 2002. *La Teoría de la Dependencia: Balance y Perspectivas*. Buenos Aires: Plaza and Janes. Takes stock of dependency theory and suggests that globalization has reinforced the types of problems that were originally raised in the 1960s.

Eichengreen, Barry, Ricardo Hausmann, and Ugo Panizza. 2003. "Currency Mismatches, Debt Intolerance, and Original Sin." NBER Working Paper No. 10036. Cambridge, MA: National Bureau of Economic Re search. Discusses the volatility of emerging market economies and the difficulty these countries have in servicing and repaying their debts.

Frank, André Gunder. 1969. *Capitalism and Under development in Latin America: Historical Studies of Brazil and Chile*. New York: Monthly Review. The closest to a dependency theory manifesto.

Tavares, Maria da Conceição. 2000. "Subdesenvolvimento, Dominação, e Luta de Classes." In *Celso Furtado e o Brasil*, edited by Maria da Conceição Tavares. São Paulo: Fundação Perseu Abramo. Suggests that international finance, and not technology, is at the center of the de pendency situation.

Topik, Stephen. 1998. "Dependency Revisited: Saving the Baby from the Bathwater." *Latin American Perspectives* 25 (6) (November): 95 99. Provides a review of the dependency literature that emphasizes its connection to other schools of thought.

Vernengo, Matías. 2006. "Technology, Finance, and De pendency: Latin American Radical Political Economy in Retrospect." *Review of Radical Political Economics* 38 (4) (fall): 551 68. A critical assessment of the debates about dependency in Latin America.

Wallerstein, Immanuel. 1976. *The Modern World System: Capitalist Agriculture and the Origins of the Eu ropean World Economy in the Sixteenth Century*. New York: Academic Press. The first of a three volume series that describes the rise of the world system, a complex network of exchanges and production on a global scale.

MATÍAS VERNENGO

■ deposit insurance

Deposit insurance is part of the "safety net" for the banking system, but also can contribute to the system's instability, depending on its coverage and design. Deposit insurance has cross-border implications through its impact on banking crises that potentially spread to other nations.

In most countries, the regulatory structure for banks is much farther reaching than regulation of nonfinancial firms as well as of nonbank financial

institutions. The case for regulation is based on special characteristics of banks and perceived market failures in banking. First, the double role of banks as liquidity providers and participants in credit and capital markets makes them potentially vulnerable to bank runs, since a large share of the assets cannot be liquidated quickly in case depositors want to convert their funds to cash. Second, banks are relatively opaque entities, making it difficult for depositors and other creditors to evaluate the default risk of each bank. The limited information among depositors about the risk and value of bank assets can lead to the spread (contagion) of bank runs from one bank to another (Diamond and Dybvig 1983). Third, there are generally substantial amounts of very short short-term interbank liabilities, which contribute further to the risk of contagion. The potential for contagion implies that the banking system is subject to "systemic risk" to a greater extent than other providers of credit. Fourth, banks play a key role in payment and settlement systems, with the implications that large failures can disrupt economic activity.

The risk of runs on a bank and contagion implies that speed of action is of the essence when a bank is perceived to be near failure. Conventional liquidation and restructuring procedures for corporations are too time-consuming to be applied to banks without modification.

Some economists argue that banks are not qualitatively different from other firms and that markets provide sufficient discipline on banks' risk taking. These economists, who generally are proponents of "free banking," are few, however. From a policy point of view it is most important which view dominates among policymakers. It can safely be said that in times of crisis they will not be willing to experiment in order to find out which group of economists is right.

The need for rapid intervention, lack of effective rules for dealing with a bank in distress, and fear of contagion work in tandem to compel governments to intervene in a crisis by issuing blanket guarantees to all creditors or bailing most or all of them out. Anticipating this government behavior, depositors,

other creditors, and sometimes shareholders as well perceive themselves as implicitly insured even if there is no explicit deposit insurance.

The Safety Net for the Banking System The features of banking described earlier have led to the implementation of a number of measures in most countries that together constitute the safety net for the banking system. The typical components of the safety net are:

1. A lender of last resort (LOLR)
2. Deposit insurance
3. Supervision and regulation of banks' risk taking
4. Capital requirements

These aspects of the safety net should jointly protect the "safety and soundness" of the banking system while providing banks with the appropriate rules and incentives to allocate credit efficiently. Deposit insurance can limit the risk of bank runs by guaranteeing that depositors receive some, or all, of their deposited funds with reasonable speed in case their banks become insolvent or illiquid. The central bank can also act as a LOLR by lending to a solvent bank facing a liquidity squeeze as a result of a run by depositors. To limit the LOLR to cases of illiquidity, the central bank can require collateral to provide liquidity support.

Regulation of banks' behavior and asset allocation, and supervision of banks' credit allocation and risk management systems, have the purpose of limiting banks' risk taking, which can be excessive. In addition, capital requirements reduce risk-taking incentives by ensuring that there is always shareholder capital at risk. Capital also serves as a buffer against unanticipated losses.

International agreements with respect to capital requirements as well as principles for supervision are negotiated within the so-called Basel Committee. The first Basel Agreement was completed in 1988 (Basel I). After years of debate, a substantially revised Capital Adequacy Accord (Basel II) was completed in 2004. This accord was to be implemented in the European Union (EU) in 2007 and 2008, but it remains controversial, particularly in the United States.

Deposit Insurance, Risk Taking, and Banking Crises The flip side of the positive role of deposit insurance as a safeguard against bank runs and as a consumer protection device is its role in inducing banks to shift risk to a deposit insurance fund or tax payers. These risk-shifting incentives are caused by limited liability of shareholders and explicit or implicit protection of depositors and other creditors. The so called moral hazard problem caused by these factors implies that banks have incentives to take on excessive risk on the asset side or to keep the equity capital low. Thus deposit insurance systems can contribute to the very problem (systemic risk) they are designed to reduce.

One solution to the moral hazard problem would be to design a deposit insurance premium structure reflecting banks' risk taking. Risk-based pricing encounters the problem of defining banks' risk taking contractually. For this reason a private deposit insurance market is not likely to function well. The existence of explicit and implicit insurance also undermines the scope for private insurance. In the United States, the Federal Deposit Insurance Corporation (FDIC) sets insurance premiums based on levels of capital.

The substantial resources devoted to the design of a capital adequacy framework by central bankers and regulators in the Basel Committee indicate that there is a strong concern about incentives for excessive risk taking. Bank managers, by contrast, tend to deny that there are incentives for excessive risk taking because they do not deliberately set out to take "excessive" risk. Incentives need not reveal themselves as incentives for deliberate risk taking, however. Instead, it is the competition among banks with the opportunity to finance their lending activities at a near risk-free interest rate that induces them to prefer debt financing to equity financing. Furthermore, competition for funding will not be based on banks' risk evaluation and risk management skills. Increased resources devoted to regulation and supervision and increased sophistication of supervisors have done little to reduce the incidence of banking crises. For this reason several academic economists have called for increased reliance on market discipline in the regulatory framework for banks.

The existence of implicit insurance implies that it is not necessarily the extent of explicit insurance that determines creditors' and, indirectly, banks' behavior. Absence of explicit insurance does not constitute credible noninsurance if political realities require supervisors and governments to rapidly intervene in banking crises to protect creditors. On these grounds a deposit insurance scheme with limited coverage can maximize the market discipline by making it credible that noninsured deposits and creditors will not be bailed out (Angkinand and Wihlborg 2006). The appropriate coverage depends on a number of country-specific institutional factors affecting credibility of noninsurance. In particular, effective procedures for dealing with a bank in distress can reduce the likelihood of bailouts.

The empirical evidence on the relationship between the coverage of deposit insurance schemes and risk taking is ambiguous. Much work has been devoted to analysis of the relationship between deposit insurance coverage and the occurrence of banking crises around the world.

Differences in results across studies suggest that institutional differences matter greatly for the effect of deposit insurance on the likelihood of banking crisis. The economists Demirgüç-Kunt and Detragiache (2002) consider the effectiveness of prudential regulation and supervision, as well as the strength of the legal system, finding that deposit insurance contributes less to the probability of banking crisis in countries with a high level of institutional quality.

Banking crises and excess risk taking have also been analyzed at the bank level. On this level it is necessary to take into account that capital and risk taking are determined simultaneously. The economists Nier and Baumann (2006) found that bank capital is decreasing in deposit insurance coverage, increasing in uninsured deposits, and decreasing in government support. These results provide evidence that market discipline depends on explicit coverage as well as the credibility of noninsurance.

Dimensions of Deposit Insurance Systems

Several dimensions of explicit deposit insurance schemes can discourage bank runs and incentives for risk taking. In discouraging bank runs, speed and credibility of insurance compensation in case of a bank failure are particularly important.

Risk-taking incentives are influenced by the existence of groups of creditors with incentives to monitor banks and to withdraw funds if they find that a bank takes unacceptable risks. Demirgüç-Kunt and Detragiache (2002) constructed a "moral hazard index" for a large number of countries from data on coinsurance features, coverage of foreign currency and interbank deposits, type of funding, source of funding, management, membership, and the level of explicit coverage. Coinsurance implies that those insured are responsible for parts of the losses. Foreign currency and interbank deposits are not covered in most countries. Private participation in the deposit insurance system can also be required.

Funding can be through insurance premiums that are used to build up a fund, or governments can cover payments out of tax revenues when compensation is due. The existence of a fund enhances the credibility of the system and it implies that banks pay insurance premiums, which can be based on proxies for risk. Most countries either do not charge a premium or they charge a certain percentage of deposits. The method to replenish losses to a fund, and assign responsibility in case the fund is insufficient, also affects incentives. If banks are held responsible for the replenishment of a fund, the subsidy component to the insurance system is reduced and banks have stronger incentives to watch the soundness of the system. Finally, membership in the system can be voluntary or required.

The economists Hovakimian, Kane, and Laeven (2003) have estimated implicit insurance premiums for banks in a number of countries based on the insight that deposit insurance can be interpreted and valued as a put option on banks' assets (i.e. the right but not obligation to sell). The authors find that some risk-adjustment of the insurance premium for each bank, the existence of coinsurance, and the funding of the deposit insurance along with deposit insurance coverage affect the variation in implicit insurance premiums across countries.

Deposit Insurance in Cross-Border Banking

Banks are involved in cross-border activities through direct lending to foreign banks, companies, and governments; through subsidiaries; and through branches operating in foreign countries. In the first case the domestic deposit insurance system affects the incentives for lending to foreign entities and protects domestic depositors. Implicit insurance can also be provided by international financial institutions, in particular the International Monetary Fund (IMF). The IMF's role in helping Mexico during the crisis in 1995 has been widely blamed for contributing to the Asian crisis in 1997. The IMF does not protect banks directly but by helping countries such as Mexico in balance of payments crises, it indirectly provides protection for banks that have lent to banks and the government in a crisis country. As a result banks may consider loans to foreign governments and banks protected by the same governments to be relatively safe.

Foreign subsidiaries are separate legal entities. Depositors in subsidiaries are therefore protected by host country deposit insurance systems. Host countries are also responsible for supervision of the subsidiaries while home countries are responsible for supervision of the consolidated bank. Since subsidiary operations often are closely integrated with the parents in complex financial organizations, there are opportunities for shifting of risk between the entities. If a subsidiary or the consolidated bank fails there is scope for conflicts of interest between host and home countries with respect to the sharing of the burden for losses that are not clearly attributable to one of the entities. The absence of predetermined procedures for crisis resolution in banks makes conflicts more likely. It is common for home and host countries to have memoranda of understanding on crisis resolution for banks, but these memoranda are typically so general that they offer little guidance if a bank fails.

A bank can also run its host country operations through branches. Since branches are not separate legal entities, the challenges for authorities in home and host countries are different. The EU's Banking Directive provides banks with the opportunity to operate across borders within the EU through branches under home country supervision. Deposit insurance in the EU is also a home country responsibility. Thus banks of different nationalities can operate in the same country offering different levels of deposit insurance. If a host country has a higher coverage, local branches of foreign banks can be offered the opportunity to "top up" their deposit insurance. The EU rules with respect to deposit insurance and supervision are consistent in the sense that the responsibilities for insurance and supervision coincide. Nevertheless, there is substantial worry in host countries that their interests will not be well represented if a foreign bank fails. These concerns partly explain why the Banking Directive has not been put into large-scale practice; instead, most cross-border banking is organized in subsidiaries.

The situation in the United States is different. Branches of foreign banks must participate in the U.S. deposit insurance system, although they are foreign legal entities and formally relying on the same capital as buffer against losses. To protect the American interests in case a foreign bank fails, branches of foreign banks are "ring-fenced," meaning that the branches must hold separate capital for their activities in the United States. Therefore, the difference between a subsidiary and a branch in the United States is almost in name only.

See also asymmetric information; bailouts; contagion; discipline; financial services; International Monetary Fund (IMF); lender of last resort; spillovers

FURTHER READING

Angkinand, Apanard, and Clas Wihlborg. 2006. "Bank Insolvency Procedures as Foundation for Market Discipline." In *Cross border Banking* (Proceedings from Federal Reserve Bank of Chicago The World Bank Conference), edited by G. Caprio, D. Evanoff, and G. Kaufman. Chicago: World Economic Publishers, 423 44. Shows how a deposit insurance scheme with limited coverage can maximize the market discipline.

Barth, James R., Gerard Caprio Jr., and Ross Levine. 2006. *Rethinking Bank Regulation: Till Angels Govern*. Cambridge: Cambridge University Press. Provides an analysis of the effectiveness of banking regulation and supervision in many dimensions.

Benink, Harald, and George Benston. 2005. "The Future of Banking Regulation in Developed Countries: Lessons from and for Europe." *Financial Markets, Institutions, and Instruments* 14 (5): 289 328. Shows how banks' equity capital relative to total assets worldwide declined from a level similar to nonfinancial firms in the 1920s to a level of around 4 percent in the late 1980s when the Basel Committee began its work.

Benston, G. J., R. A. Eisenbeis, P. M. Horvitz, E. J. Kane, and G. G. Kaufman. 1986. *Perspectives on Safe and Sound Banking*. Cambridge, MA: MIT Press. Includes articles by several authors about different aspects of the safety net for banks.

Bhattacharya, Sudipto, and Anjan V. Thakor. 1993. "Contemporary Banking Theory." *Journal of Financial Intermediation* 3 (1): 2 50. Elaborates on the theory underlying principles for banking regulation and supervision.

Caprio, Jerry, and Daniela Klingebiel. 2002. "Episodes of Systemic and Borderline Financial Crises." Discussion Paper No. 428. Washington, DC: World Bank. Provides evidence of excess risk taking in the frequency of banking crises around the world. The documentation in this article is used in most empirical work on banking crises.

Demirgüç Kunt, Asli, and Enrica Detragiache. 2002. "Does Deposit Insurance Increase Banking System Stability? An Empirical Investigation." *Journal of Monetary Economics* 49 (7): 1373 1406. Empirically analyzes the relationship between deposit insurance and banking crises.

Demirgüç Kunt, Asli, B. Karacaovali, and Luc Laeven. 2005. "Deposit Insurance around the World: A Comprehensive Database." World Bank Policy Research Working Paper No. 3628. Washington, DC: World Bank. Database described in this article is used in most of the empirical work on deposit insurance.

Diamond, Douglas W., and Philip H. Dybvig. 1983. "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy* 91 (3): 401 19. Classic article, which

develops a theoretical model of bank runs and liquidity crisis.

Hayek, Friedrich. 1977. *Denationalisation of Money*. London: Institute for International Economics. Provides an argument for "free banking" without central bank monopoly on issuing money.

Hovakimian, Armen, Edward Kane, and Luc Laeven, 2003. "How Country and Safety Net Characteristics Affect Bank Risk shifting." *Journal of Financial Services Research* 23 (3): 177 204. Estimates implicit deposit insurance premiums for banks in a large number of countries and analyzes how these premiums depend on characteristics of deposit insurance systems.

Kane, Edward J. 2000. "Designing Financial Safety Nets to Fit Country Circumstances." Policy Research Working Paper 2453. Washington DC: World Bank. Institutional features of deposit insurance systems and other aspects of the safety net are documented and discussed in this article.

Nier, Erlend, and Ursel Baumann. 2006. "Market Discipline, Disclosure, and Moral Hazard in Banking." *Journal of Financial Intermediation* 15 (3): 332 61. Examines the links between bank capital and deposit insurance coverage.

CLAS WIHLBORG

■ development

Development refers to a process that includes and goes beyond economic growth. Whereas growth is defined narrowly as an increase in real income per person, the Nobel laureate Amartya Sen (1999) defines development as the expansion of human freedom, including freedom from hunger, ignorance, political oppression, and disease. Although income clearly matters for development a richer society can build and staff more schools, courtrooms, and hospitals it is also clearly not all that matters.

In defining development the challenge is to find measures of social evolution that come closer to Sen's notion of expanding human freedom but are still concrete enough to permit meaningful measurement and study. Making choices about which social goals are emphasized and how their attainment is mea-

sured is of necessity a value-laden exercise, and because values vary sharply across cultures and individuals, no measure of development can safely claim to be either universal or entirely objective. Despite this inherent limitation, the inadequacy of economic growth as a measure of development has led to an ongoing effort to find alternative measures of development. Here we consider several dimensions of development economic, human, and social noting how each has been defined and measured and considering the relationship of development to economic growth.

Economic Development Economic development includes three dimensions that go beyond rising per capita incomes. The first involves the development of a country's economic system. Economic development is generally facilitated by a number of structural changes, including urbanization, the rise in the size of firms, the relative decline of the agricultural sector in terms of employment and output with expansion of manufacturing and services, the geographic expansion of markets, and increases in the diversity of goods produced and traded. These changes in the organization of economic activity are the hallmark of permanent shifts in economic activity. Income gains that are not accompanied by these changes, such as natural resource booms, are easily reversed. Furthermore, the expansion of markets and increases in the diversity of goods produced raises welfare by expanding the range of goods available to consumers and providing a form of insurance by diminishing the impact of dramatic changes to the local economy or particular industries.

The second dimension of economic development concerns the distribution of the gains from economic growth. A rise in the average income level means nothing more or less than a rise in national income divided by population size. As such, there is no guarantee that economic growth makes a majority of people better off or results in income gains for the poorest members of society. This point is particularly important for less-developed countries, in which calls for economic growth are motivated largely by the desire to reduce widespread poverty.

The perception that the gains from growth in the 1960s and 1970s were unequally distributed led to calls to redefine development in a manner that placed greater emphasis on poverty and inequality (see, e.g., Streeten 1981). Skepticism about the gains from growth also reflected the influence of economists such as the Nobel laureates W. Arthur Lewis and Simon Kuznets, who argued that industrialization was characterized by an initial period of rising inequality with limited benefits for the poor. In addition, many economists noted that the way in which growth rates are calculated gives greater weight to increases in the incomes of rich individuals: a 10 percent increase in an individual's income adds more to average income the richer that person is. One suggestion was to replace growth with a poverty-weighted index that measured the growth of the incomes of the poorest 20 to 40 percent of the population.

Although development economists remain concerned with the distribution of the gains from growth, the idea that growth is inherently biased against the poor is not supported by recent empirical work. For example, a study by Dollar and Kraay (2002) finds that on average the incomes of the poorest 20 percent of the population grow at the same rate as average income. Similarly, in a cross-country analysis of income inequality, Li, Squire, and Zou (1998) find that inequality is highly stable within countries over the postwar period. Prominent exceptions include rising inequality in the United States and falling inequality in several fast-growing East Asian economies. Although these results do not guarantee that the poor always benefit from growth, they undermine the presumption that growth is systematically biased against the poor.

The final dimension of economic development is sustainability (see, e.g., Goldin and Winters 1995). The UN's Brundtland Commission (1987) defined sustainable development as development that "meets the needs of the present generation without compromising the ability of future generations to meet their own needs." Viewed in terms of sustainability, income growth rates can be misleading measures of development, since consuming natural resources simultaneously increases income and reduces a country's stock of natural wealth. Correctly measuring sustainable development requires adjusting income levels to account for the depletion of natural resources and degradation of environmental quality. This has led to the development of "green" national accounting, in which the national saving rate is adjusted for resource depletion. Adjusting income levels for the negative impact of pollution is more difficult, since the value of environmental quality varies widely across individuals, societies, and time.

The relationship between economic growth and environmental quality is not clear cut. Economic growth clearly increases the pressure on natural resources and systems. The challenges to sustainability include the consumption of nonrenewable resources such as oil and coal reserves, loss of biodiversity, depletion of ocean fisheries, deforestation and desertification, and reductions in air and water quality, all of which reduce either the productivity or consumption of future generations. But rising income levels also increase a society's willingness to pay for environmental quality, its ability to mobilize political pressure, and the technologies and resources it has available to reduce or reverse environmental damage. Together these forces result in complex relationships between environmental quality and average income levels. For example, as income levels rise, access to clean water increases, carbon dioxide emissions worsen, and the output of industrial pollutants such as sulfur dioxide rises and then falls.

Human Development If income matters for development, it is because it measures the ability of individuals to satisfy their needs and create personally fulfilling lives. In brief, a higher income is at best a means to the end of greater development. Recognizing this led to several efforts to measure the "ends" directly, resulting in attempts to quantify numerous dimensions of consumption, including consumption of food, medical care, housing, and energy. Many of these measures were subsequently criticized as being based on Western norms and consumption patterns   larger houses counted but larger weddings did not   reducing their usefulness as general measures of development. A more gener-

ally accepted approach is the Human Development Index, which has been published annually since 1990 as part of the UN's Human Development Report.

The Human Development Index, or HDI, gives equal weight to three components measuring income, health, and educational attainment. For each component, a country's outcomes are graded on a scale from 0 to 1, where the ends of the spectrum are defined by the lowest and highest observation for other countries. Health is measured by life expectancy at birth. Life expectancy is highly sensitive to infant and child mortality rates. Since these depend on widespread access to nutrition and health care, life expectancy is highly sensitive to economic and social inequality. Education is a fundamental determinant of economic and social mobility, and as such plays an important role in determining the range of choices available to an individual. The educational index is calculated as two-thirds adult literacy and one-third current enrollment rates, making it sensitive to both educational inequality and future educational attainment. The HDI includes a measure of income to proxy for all aspects of human development not captured by health and education. The index of income levels uses the natural log of per capita income. This choice places greater weight on income increases among poor nations, since it implies that proportional increases in income are measured equally: the increase from $1,000 to $2,000 is equal to that from $10,000 to $20,000.

Although it is still clearly limited, the HDI is the most widely known and commonly used alternative to per capita income as a measure of development. To see how the HDI differs from per capita income as a measure of development, it is useful to compare country rankings using the HDI and per capita income levels. Using data from the 2007 Human Development Report, we find that major oil exporters and countries with significant AIDS crises do worse when ranked according to human development, results that probably reflect the HDI's sensitivity to inequality and life expectancy. Similarly, the United States ranks lower using HDI than it does using per capita income, an outcome that reflects inequalities in access to education and health care

relative to other wealthy countries. Alternatively, current and formerly socialist countries, such as Cuba, Venezuela, and the former Soviet countries, tend to rank higher using the HDI, reflecting both their commitment to the provision of education and health care and their relative inability to generate high incomes for their citizens. These differences aside, differences in per capita income explain more than 88 percent of the variation in the HDI, suggesting that income is the most important determinant of both health and educational outcomes. With higher incomes societies can build schools and hospitals and hire the teachers and doctors to staff them.

Social Development Social development refers to development of society as a whole. A number of dimensions of social development have been measured and subjected to analysis, often as part of the UN's Human Development Report, including political development, human rights, human security, and gender equality. Here we consider two aspects of social development: political development and gender equality. It should be noted that in attempting to define social development, it is impossible to escape from subjective value judgments that vary widely across individuals and cultures. In particular, the measures presented here have their roots in Western Enlightenment thinking that is both secular and individualistic. To the degree that other cultures emphasize competing values based on family or community or religious devotion, these measures may be inappropriate.

We measure political development along two dimensions, the extension of democratic political rights and freedom from corruption. Democratic political rights may be viewed both as an end in themselves and as the means to achieve other development goals. For example, democratic political rights and the civil rights that sustain them, such as the freedoms of speech, association, and the press, are central to self-expression and to full participation in the social life of a country. They may also play an instrumental role in reducing income inequality and raising incomes by securing access to health care and education. As Sen (1981) famously pointed out, despite instances of widespread crop failure, there has

never been a famine in a democratic country: letting too many people starve is bad for reelection. Though it is difficult to measure democracy objectively, there appears to be a strong relationship between average income and democratic political rights. It is hard to disentangle cause from effect, and there appear to be important links in both directions. For example, democracies tend to have higher levels of education but also higher taxes, with conflicting implications for economic growth, and the demand for political rights may increase as rising incomes satisfy more basic physical needs (see, e.g., Barro 1996).

While democracy focuses on how political leaders are chosen, governance refers to their behavior once in office. Democracy may be seen as a means to the end of good governance. An important aspect of governance is corruption, defined as using public office for private gain. Measures of corruption are constructed using survey data and tend to be highly correlated with one another and with other measures of governance such as the quality of bureaucracy. Freedom from corruption tends to rise with per capita income, though less strongly than the Human Development Index. For example, using data from 2007, the correlation between the natural log of per capita income and Transparency International's Corruption Perceptions Index is 0.63. As with democracy, causality appears to run both ways. Clean government results in the better use of tax revenues and provision of public goods, and higher incomes increase the demand for good government and the ability to supply it.

Gender equality involves equal access for women and girls to economic opportunities and resources and full participation in the social and political spheres of life. Gender equality was first included in the Human Development Report in 1995. The Human Development Report now reports a Gender-related Development Index (GDI), which measures the inequality of health, education, and income outcomes by gender using the same indexes as the Human Development Index. Gender equality is both a fundamental measure of women's rights and opportunities and an important contributor in securing other development objectives. As with political development, there are complex relationships between gender equality and political and economic development. Expanding women's political participation may reduce corruption and increase the provision of education and health care, while women's economic participation both raises incomes directly and has important impacts on the well-being of future generations.

See also aid, international; economic development; evolution of development thinking; poverty, global

FURTHER READING

Barro, Robert J. 1996. "Democracy and Growth." *Journal of Economic Growth* 1: 1 27. Presents evidence on the re lationship between economic and political development.

Dollar, David, and Aart Kraay. 2002. "Growth Is Good for the Poor." *Journal of Economic Growth* 7 (3): 195 225. Presents evidence on the growth of incomes among the poor.

Goldin, Ian, and Alan Winters, eds. 1995. *The Economics of Sustainable Development.* Cambridge: Cambridge Uni versity Press. An excellent introduction to sustainable development.

Li, Hongyi, Lyn Squire, and Heng fu Zou. 1998. "Ex plaining International and Intertemporal Variations in Income Inequality." *Economic Journal* 108: 26 43. Ar gues that for most countries income inequality has been relatively constant in the post World War II era.

Sen, Amartya. 1981. *Poverty and Famines: An Essay on En titlement and Deprivation.* Oxford: Oxford University Press. A foundational analysis of famine in which Sen argues for the importance of democratic political rights.

. 1999. *Development as Freedom.* New York: Knopf. An extended essay on development by the foremost thinker in this area.

Streeten, Paul. 1981. *First Things First: Meeting Basic Needs in Developing Countries.* Oxford: Oxford University Press. Argues that development should be defined in terms of poverty reduction rather than growth.

United Nations. 1987. "Report of the World Commission on Environment and Development." General Assembly Resolution 42/187. New York: UN. Known as the Brundtland Commission report, this paper did much to popularize the idea of sustainable development.

LEWIS S. DAVIS

■ digital divide

The term *digital divide* refers to the disparity between those who have the resources and capabilities to use telecommunications technologies and those who don't. This gap exists within nations (varying by geography, social class, age, and other social dimensions) and between nations (essentially reflecting more fundamental differences in social and economic development). Blending telecommunications with computer technology, the advent of the Internet constitutes one of the latest revolutions in how the world conducts its business, thereby making the digital divide ever more salient. Evolving from modest beginnings as a U.S. Defense Department project in the 1960s to a mass communications technology linking the world's cities, the Internet/World Wide Web has quickly outgrown its original role as a special tool of governmental and educational elites. As such, this new technology is simultaneously hailed as the fifth neo-Schumpeterian "long wave" of global economic expansion (Castellacci 2006) and cursed as a new, powerful tool of exploitation created by international capitalism (Parayil 2005). This debate, between those Norris (2001) labels as Internet "optimists" and "pessimists," animates much of the scholarly literature, but the truth about digital diffusion is probably more complicated (and less economically crucial) than many analysts suggest.

There is no doubt that the debate is driven by a real phenomenon. Although the number of Internet users has expanded exponentially from approximately 40 million in 1995 to more than a billion people in 2007, the Western world's share of the total number of users worldwide is still around 50 percent, even though the West makes up only about 17 percent of the world's population. On the other hand, the combined share of users from Africa, the Middle East, and Latin America (about a quarter of the globe's population) constitutes only about 13 percent of all users, suggesting a considerable disparity. Even more telling, as of 2007 about 70 percent of the populations of the United States and Canada used the Internet, while only 11 percent and 3 percent of Asia's and Africa's populations, respectively, did the same (Internet World Stats 2007).

For all intents and purposes, then, the Internet/World Wide Web is still very much a creature of well-developed markets. Although optimists hope that developing countries can use the Internet to "leapfrog" over the industrial phase of development and move directly into postindustrialism (Steinmueller 2001), the commodification of information is sufficiently different from material goods and other services to cast doubt on this proposition. Many goods and services are produced for final consumption, whereas information is typically produced as an intermediate commodity, something used to improve the efficiency of production or the distribution of goods and services, much like transportation (Mowshowitz 1992; Crenshaw and Robison 2006a). This explains why the need for information grows as social and economic complexity increases, and thus why demand for information as a commodity is contingent on large and complex markets.

Since it is clear that the Internet is firmly rooted in developed economies and is likely to remain so, it is the *diffusion* of Internet technology to the developing world that dominates the current scientific literature on the digital divide. Rogers (1983) defines cultural diffusion as "an innovation that is communicated through certain channels over time to a receiving society." Not surprisingly, the current literature focuses almost exclusively on these "receiving societies," attempting to gauge their *structural conduciveness* as adopters of complex information and communications technologies (ICT). The most consistent structural attributes that have been found conducive to Internet diffusion fall into a modest number of categories: (1) economic complexity, infrastructure, and income; (2) mass education and literacy; (3) public policies relating to the provision of telecommunication services; and (4) the relationship between government and civil society, with special emphasis on political rights and property rights. In general, higher incomes, a dominant services sector, more extensive telephone and electrical infrastructure, mass education and literacy, public policies related to telecommunications competition, reasonably priced telecommunications services, participatory democracy, and protection of property rights

correlate positively with more rapid adoption of the Internet.

Diffusion theory stresses *contact* in addition to conduciveness, however. Indeed, the major reason Internet optimists think that "leapfrogging" is possible is that the Internet makes the world's most complicated economies accessible to a wide range of less-developed communities, even those that are relatively isolated. The thinking suggests that the Internet provides conduits that allow the populations of relatively undeveloped countries to access a complex global economy. In short, the optimistic view of Internet diffusion is predicated on a fairly benign view of globalization.

Globalization typically refers to a plethora of international networks composed of intergovernmental organizations, governments, multinational corporations, and many other actors. Although limited globalization has long been a part of human history, contemporary globalization is accomplished through unprecedented technological capacities, flows of material and information, and growth in international governance. Given this dynamism, it seems plausible to assume that globalization might "jump-start" Internet development in otherwise unlikely locations. The question is, therefore, whether globalization can ease the constraints placed on Internet diffusion by structural requisites (i.e., conduciveness).

Given that the Internet was created in the postindustrial West, contact with the West (and/or other advanced economies) may form postindustrial "bridges" or "beachheads" in developing countries (the choice of term depends on one's view of global capitalism). One interesting analogy is Crosby's (1986) notion that European colonialism was far more successful where local conditions allowed the formation of "neo-Europes" territories where climates and disease regimes were conducive to the importation of European flora and fauna, which in turn displaced many native species and invited European settlement. The bridges (or beachheads) provided by the conduits of globalization may be creating something similar for ICT growth islands of postindustrialism amid seas of the preindustrial.

Essentially, globalization creates postindustrial microcosms that generate *demand* for ICT where it would not otherwise exist.

The possible conduits of globalization should be familiar to students of international development. For instance, megacities (i.e., cities populated by millions of people) have proven to be important gateways for ICT diffusion (Crenshaw and Robison 2006a, 2006b). Urban areas generate extremely diverse labor and consumer markets and, as major service nodes, tend to host Internet service providers (ISPs) even in very poor nations. For instance, Nigeria is one of the world's poorer nations, and yet its capital of Lagos (a city of approximately 11 million people) had more than 90 ISPs in 2003. Thus such megacities can provide "islands" of postindustrial demand independent of a nation's other structural characteristics.

International investment and trade also provide conduits that create favorable environments for ICT development. One of the major reasons that a digital divide exists among developing nations is that trade and foreign investment are likewise unequally distributed. For instance, in the early 21st century, sub-Saharan Africa accounts for only 2 percent of the world's merchandising exports, and receives only about 8 percent of the foreign investment going to developing economies. If ICT diffusion tends to follow pathways blazed by trade and investment (Crenshaw and Robison 2006a), inequalities in Internet capacity and growth are probably inevitable.

While large commercial cities and international investment and trade probably play dominant roles in the spread of Internet traffic in the developing world, noncommercial/nonstate actors and flows of immigrants, some permanent and some temporary (i.e., tourists), also contribute to the diffusion of ICT. For instance, nongovernmental organizations (NGOs) and international nongovernmental organizations (INGOs) are carving ever-expanding niches in the world's cross-cultural networks. Indeed, the Union of International Associations estimates that around 70,000 NGOs and INGOs existed as of the mid-1990s (UIA 2008), and regardless of the myriad purposes of such organizations, the one thing

they have in common is dependence on international telecommunications to spread their messages and coordinate their activities. As past research demonstrates, in all likelihood this growing web of nonstate organizations makes a substantial contribution to Internet connectivity in otherwise unlikely places (Drori and Jang 2003; Crenshaw and Robison 2006a).

Tourism also promotes Internet use in the developing world. Tourist destinations that cater to affluent visitors create nodes of strong demand for instant and up-to-date information and real-time communications, not only for the tourists themselves but also for the global travel industry. As with other global connections, however, global tourism is also unevenly distributed. For instance, American tourists disproportionately travel to the Caribbean and Europe, whereas Africa and many other destinations languish in relative obscurity. Western tourism has been found to promote Internet usage (Crenshaw and Robison 2006a, 2006b), and there is very little reason to expect this dynamic to equalize Internet capacity in the future.

On balance, then, what can be said is that structural conduciveness in the form of affluence, democracy, and political and property rights does invite the spread of digital technology. Moreover, other forces, particularly economic, political, and social globalization, contribute to the digital revolution, even in places where social structural conditions may be unsuitable (e.g., poor nondemocratic countries). None of these forces, however, easily lend themselves to policies that might "jump-start" the Internet revolution, nor does the current literature suggest that Internet development is crucial for economic growth in the developing world. Rather, the global digital network is an epiphenomenon of more fundamental economic, political, and social forces, and policies *focused* on Internet development are likely putting the cart before the horse. New technologies in laptop computers and wireless connectivity will ease the supply problem constraining Internet development, but solutions to the bedrock problem underlying the digital divide — the lack of demand that results from national poverty and isolation from the world

economy — also depend on political reform, domestic production, economic trade, and tourism.

See also information and communications technology

FURTHER READING

Castellacci, Fulvio. 2006. "Innovation, Diffusion, and Catching Up in the Fifth Long Wave." *Futures* 38: 841 63. A discussion of how technological "waves" and global institutions are likely to influence the "digital divide."

Crenshaw, Edward M., and Kristopher K. Robison. 2006a. "Globalization and the Digital Divide: The Roles of Structural Conduciveness and Global Con nection in Internet Diffusion." *Social Science Quarterly* 87: 190 207. An empirical, cross national analysis of macro social factors influencing the adoption of the Internet.

———. 2006b. "Jump Starting the Internet Revolution: How Structural Conduciveness and Global Connec tions Help Diffuse the Internet." *Journal of the Associa tion for Information Systems* 7 (1): 4 18. A cross national pooled time series of the macro social factors shaping the spread of the Internet.

Crosby, Alfred W. 1986. *Ecological Imperialism: The Bio logical Expansion of Europe, 900 1900.* Cambridge: Cambridge University Press. A detailed treatise on how biological/climatic compatibilities aided European col onization of non European areas.

Drori, Gili S., and Yong Suk Jang. 2003. "The Global Digital Divide: A Sociological Assessment of Trends and Causes." *Social Science Computer Review* 21: 144 61. Research demonstrating that cultural commitment to science and education may be paramount in IT diffusion.

Internet World Stats. 2007. http://www.internetworldstats .com/stats.htm. A website that provides a variety of macrolevel statistics for most nations of the world.

Kiiski, Sampsa, and Matti Pohjola. 2002. "Cross Country Diffusion of the Internet." *Information Economics and Policy* 14: 297 310. An empirical, cross national anal ysis of internal factors that shape IT development.

Mowshowitz, Abbe. 1992. "On the Market Value of In formation Commodities, I: The Nature of Information and Information Commodities." *Journal of the American Society for Information Science* 43: 225 32. A discussion

of how and why the production and use of information is uniquely unlike other forms of economic activity.

Norris, Pippa. 2001. *Digital Divide: Civic Engagement, In formation Poverty, and the Internet Worldwide*. Cam bridge: Cambridge University Press. A treatise on the extent of the digital divide and how it may shape the economic and social future.

Parayil, Govindan. 2005. "The Digital Divide and In creasing Returns: Contradictions of Informational Cap italism." *The Information Society* 21: 41 51. An article emphasizing the "down side" of IT diffusion.

Robison, Kristopher K., and Edward M. Crenshaw. 2002. "Post industrial Trans Formations and Cyber Space: A Cross National Analysis of Internet Development." *Social Science Research* 31: 334 63. An empirical, cross national analysis of intranational forces that shape IT diffusion.

Rogers, Everett. 1983. *Diffusion of Innovations*. New York: Free Press. The "Bible" of diffusion theory.

Steinmueller, W. Edward. 2001. "ICTs and the Possibilities for Leapfrogging by Developing Countries." *Interna tional Labour Review* 140: 193 210. A discussion of the possibilities for less developed countries to use ICTs to "skip" stages of economic development.

UIA (Union of International Associations). 2008. http:// www.uia.be. A relatively comprehensive registry of nongovernmental associations around the world.

EDWARD CRENSHAW

■ **discipline**

There is a long history of arguments for the use of a fixed exchange rate regime as a source of discipline over domestic monetary and fiscal policies. Advocates of the view generally believed that governments had a tendency to debase their currencies (by generating inflation) as a way of financing their often excessive expenditures. Supporters of "sound money" were often opposed to the issuance of paper money for fear that it would lead to undisciplined money creation and hence inflation.

In the first decades of the post World War II period, such arguments were made primarily by advocates of a return to a gold standard. Mainstream Keynesian economists claimed that these arguments reflected a conservative concern with inflation to the exclusion of other macroeconomic objectives. Indeed the mainstream view was that international monetary regimes, to the extent that they were consistent with other objectives, should be devised to minimize the constraint imposed on domestic macroeconomic policymaking.

By the 1960s, however, this mainstream consensus began to fray. The combination of unemployment and inflation ("stagflation") that developed in most industrial countries during the 1960s and 1970s prompted many to question the Keynesian premise that government policies would be operated predominantly in the public interest to help correct macroeconomic instabilities generated by the private sector. Often government policies seemed to be the cause of macroeconomic instability. Although politicians were quick to adopt the Keynesian perspective that favored cutting taxes and increasing government expenditures during a recession, they were much less interested in the other side of the strategy: that taxes be raised and expenditures be cut when the economy was overheating.

Such lack of government discipline provided fertile ground for a new generation of political scientists and economists who explored why governments might deviate from the pursuit of macroeconomic stability. One influential line of thinking argued that reelection objectives could generate incentives for political business cycles (see Drazen 2000; Keil and Willett 2004). Prices tend to adjust more slowly to changes in macroeconomic policies than does output. Thus a well-timed expansion before an election might boost employment and output in the short term but lead to higher inflation in the long term. On the other hand, restrictive policies to reduce inflation might generate a recession before most of the benefits of price stability were felt.

Theoretical developments and empirical evidence helped clarify views on the trade-offs between inflation and unemployment and the time inconsistency problems to which they can give rise. These arise from differences between short term and longer run effects that lead optimization of short run net benefits

to generate greater longer run costs. In the short run, higher inflation and lower unemployment often do rise and fall together as emphasized in traditional Keynesian economics. Over the longer run, however, these effects disappear as higher or lower inflation becomes expected. Indeed, because higher inflation tends to generate greater uncertainty, over the longer term higher inflation is likely to depress employment and economic growth. Thus concerns with limiting inflation came to be seen more as a trade-off between short-term and long-term effects than between inflation and unemployment.

The new emphasis on the importance of credible commitment suggests that in addition to providing constraints over domestic macroeconomic policies the adoption of such commitments could substantially reduce the unemployment cost of disinflation by increasing the credibility of such policies. This gave rise to the popularity of exchange-rate-based stabilization strategies.

Exchange-Rate-Based Stabilization The traditional arguments for the adoption of pegged exchange-rate regimes as a means of keeping inflation low were based on the "discipline hypothesis," which proposes that pegged rates take away a government's ability to run inflating policies, or at least increase the costs of pursuing such policies.

The credibility argument, based on the theory of rational expectations, combined with the successful disinflations in the 1970s of a number of countries belonging to the pegged-rate-based European Monetary System (EMS), contributed to the popularity of exchange-rate-based stabilization (ERBS) in the 1980s. According to the credibility hypothesis, high-inflation countries can peg their exchange rates to the currency of a low-inflation country, thus sending a signal to the public that they intend to adopt and maintain a low-inflation policy. The track record of countries carrying through anti-inflationary policies to a successful conclusion had frequently been poor, so the initiation of such a program often failed to generate initial expectations that it would be successful. Therefore, there would be little initial adjustment in inflationary expectations, and the short-run effects of restrictive macroeconomic poli-

cies would show up primarily in higher unemployment and slower economic growth, which in turn made it more difficult for the government to maintain these policies. If the government could effectively signal a stronger commitment to maintaining these policies, then inflationary expectations would fall more quickly and the effects of the restrictive macroeconomic policies would show up more in lower inflation and less in higher unemployment. This in turn would enhance the odds of success.

A major focus of the credibility hypothesis is that adopting pegged exchange rates would discipline not only policymakers, but also private agents. Rational expectations theory suggests that credible institutional commitments would also discipline actions by influencing the expectations of private agents. Although some of the earlier discipline arguments for fixed exchange rates also posited that wage discipline would be enhanced since high wage increases would be less likely to be accompanied by expansionary monetary policies and hence would be likely to generate more unemployment, the emphasis on effects on private sector behavior is the focus of the rational expectations analysis (Tavlas 2000).

In the 1980s and 1990s many countries, especially in Latin America, adopted various forms of pegged exchange rates to counteract high inflation. Many academics and International Monetary Fund (IMF) officials advocated the ERBS policy, in part because of its touted ability to rapidly decelerate inflation while generating an initial boom. The advocates of ERBS, however, rarely acknowledged that the fall in inflation was often not rapid enough to prevent overvaluation of the currency. Once the currency becomes overvalued, a country's exports become less competitive, resulting in a recession and currency crisis. For ERBS to succeed, the government adopting it must also enact noninflationary fiscal and monetary policies, and convince speculators of the program's credibility.

Some advocates of ERBS have argued that such strategies should be abandoned before the need for large nominal devaluation after an excessive real overvaluation, in order to avoid a currency crisis. The propensity of ERBS to contribute to crises can easily

be seen in Latin America. Faced with rapid rates of inflation in the 1990s, many Latin America countries, including Mexico, Chile, Brazil, and Argentina, turned to ERBS, hoping for salvation. In most cases, inflation did indeed fall substantially after the ERBS had been put in place. It typically did not fall sufficiently, however, to avoid exchange rate overvaluation and eventual currency crises.

Some economists therefore began to advocate premature exits from initial pegs without fully recognizing that the temporary nature of these pegs would likely undercut their credibility. Further analysis suggested that the same types of time inconsistency problems that created the need for discipline in the first place would also make it difficult for governments to exit from a peg in a timely manner (Willett 1998). Thus pegged exchange rates are less of a constraint on inflationary policies than genuinely fixed exchange rates and generate greater incentives to pursue those same policies.

Not only did many of the ERBS efforts of the 1980s and 1990s end in crises, but research also questioned whether the EMS countries had really been able to disinflate more effectively and at lower costs in the 1970s than other industrial countries that had adopted flexible exchange rates. We should not conclude from these experiences that ERBS never makes sense, but there is now widespread recognition that many early advocates oversold its benefits and underestimated its costs. The success of the credibility hypothesis rested on a government's ability to persuade the private sector that it would pursue sound macroeconomic policies to avoid the high exit costs associated with abandoning the peg.

Discipline from International Financial Markets In contrast, the foundation of the discipline hypothesis is coercion. In the 1970s many commentators argued that international financial markets could provide external discipline. The participants in these markets would carefully monitor national economic policies and provide early warning signals if policies started to go astray. If the warnings were ignored, the resulting capital outflows and increases in interest rates could force governments and the general public to take corrective actions. Because international financial markets may not always operate in such an efficient manner, however, some analysts argue that the discipline generated by international capital flows can be capricious rather than beneficial, and high international capital mobility may reduce rather than increase discipline, thus becoming a source of instability.

See also capital flows to developing countries; capital mobility; currency board arrangement (CBA); currency crisis; European Monetary Union; exchange rate regimes; expenditure changing and expenditure switching; financial crisis; gold standard, international; impossible trinity; inflation targeting; International Monetary Fund (IMF); monetary policy rules; seigniorage; time inconsistency problem

FURTHER READING

For discussion of political business cycles and other possible causes of undisciplined macroeconomic policy, see:

Drazen, Allan. 2000. *Political Economy in Macroeconomics.* Princeton, NJ: Princeton University Press.

Keil, Manfred, and Thomas D. Willett. 2004. "Political Business Cycles." In *Encyclopedia of Public Choice,* edited by Charles Rowley and Fredrich Schneider. New York: Springer Verlag, 2:411 15.

For discussion of the possible discipline effects of pegged exchange rates and the use of exchange rates as nominal anchors, see:

Tavlas, George S. 2000. "On the Exchange Rate as a Nominal Anchor: The Rise and Fall of the Credibility Hypothesis." *The Economic Record* 76 (233): 183 201.

Willett, Thomas D. 1998. "Credibility and Discipline Effects of Exchange Rates as Nominal Anchors: The Need to Distinguish Temporary from Permanent Pegs," *World Economy* 21 (6): 803 26.

For empirical evidence on the effects of exchange rate regimes on fiscal policy, see:

Edwards, Sebastian. 2003. "Dollarization." In *The Dollarization Debate*, edited by Dominick Salvatore, James W. Dean, and Thomas D. Willett. New York: Oxford University Press, 111 29.

Tornell, Aarón, and Andrés Velasco. 1998. "Fiscal Discipline and the Choice of a Nominal Anchor in Stabilization." *Journal of International Economics* 46 (1): 1 30.

For analysis of the effects of international capital flows on monetary and fiscal policies, see:

Tytell, Irina, and Shang Jin Wei. 2004. "Does Financial Globalization Induce Better Macroeconomic Policies?" IMF Working Paper WP/04/84. Washington, DC: International Monetary Fund.

Willett, Thomas D. 2000. "International Financial Markets as Sources of Crisis or Discipline: The Too Much, Too Late Hypothesis." Princeton Essays in International Finance No. 218. Princeton, NJ: Princeton University.

SIRATHORN DECHSAKULTHORN
AND THOMAS D. WILLETT

■ **dispute settlement**
See World Trade Organization dispute settlement

■ **dissemination risk**
See foreign direct investment and international technology transfer

■ **distortions to agricultural incentives**
The majority of the world's poorest households depend on farming for their livelihood. Two sets of policy interventions have depressed earnings from farming in many developing countries since the 1950s: own-country policies, which typically have had a prourban, antiagricultural bias, and high-income country policies, which typically assist and protect farmers with import barriers and subsidies. Governments in numerous developing countries have made considerable progress over the past two decades in reducing their own sectoral and trade policy distortions to market signals, and many of them now believe high-income countries should reduce their remaining protectionism that restricts market access for developing country exports of farm products. Indeed, developing country governments have called for such commitments on farm policies in the World Trade Organization's (WTO's) current round of multilateral trade negotiations (the

Doha Development Agenda, or Doha Round) before they consider any further reform commitments of their own.

Research reported in Anderson, Martin, and van der Mensbrugghe (2006) suggests that the agricultural protectionist policies of high-income countries harm many developing countries. That research, using the global economy wide applied general equilibrium (AGE) model known as Linkage, along with similar research using the global GTAP Model (Anderson and Valenzuela 2007), also suggests full global liberalization of merchandise trade would raise value added in agriculture (that is, net farm incomes) in most developing country regions, and by more than it would raise value added in the rest of those economies (while doing the opposite in the protective high-income countries of the northern hemisphere).

Origins of Agricultural Protection While much of the government intervention in agricultural trade over the centuries has been aimed at stabilizing domestic food prices and supplies, such policy interventions tend to change in the course of a country's development from effectively taxing agriculture relative to other tradable sectors to effectively subsidizing farmers. From the late 1100s to the 1660s, prior to the first industrial revolution, Britain used export taxes and licenses to prevent domestic food prices from rising excessively. During 1660 90 a series of acts gradually raised food import duties (making imports prohibitive under most circumstances) and reduced the export restrictions on grain, provisions that were made even more protective by the corn law of 1815. The famous repeal of the corn laws in the mid-1840s heralded a period of relatively unrestricted food trade for Britain, but agricultural protection returned there in the 1930s and has been increasing since then. In many other Western European countries, similar tendencies have been observed, although on the Continent the period of free trade in the 19th and early 20th centuries was considerably shorter, and agricultural protection levels since the early twentieth century have been somewhat higher than in Britain (Kindleberger 1975). Gulbrandsen and Lindbeck (1973) estimate that the

average nominal rate of agricultural protection in Western Europe (the proportion by which domestic prices are raised by import restrictions) increased from less than 30 percent in the 1930s through the early 1950s to around 40 percent in the late 1950s and 60 percent by the late 1960s. Meanwhile, tariffs on Western European imports of manufactured goods have progressively declined since the General Agreement on Tariffs and Trade (GATT) came into force in the late 1940s. This divergence in protection trends for farmers and industrialists in developed countries was foreshadowed, as early as 1958, as a potential problem for the GATT contracting parties by a high-level panel of experts chaired by Haberler (1958).

Japan provides an even more striking example of the increasing tendency to assist agriculture more than other industries. Its industrialization began later than in Western Europe, after the opening up of the economy following the Meiji Restoration in 1868. By the early 1900s Japan had switched from being a small net exporter of food to becoming increasingly dependent on rice imports. This was followed by calls from farmers and their supporters for rice import controls. Their calls were matched by equally vigorous calls from manufacturing and commercial groups for unrestricted food trade, since the price of rice at that time was a major determinant of real wages in the nonfarm sector. The heated debates were not unlike those that had led to the repeal of the corn laws in Britain six decades earlier. In Japan, however, the forces of protection triumphed, and a tariff was imposed on rice imports beginning in 1904. That tariff gradually rose over time, providing a nominal rate of protection for rice of more than 30 percent during World War I. Even when there were food riots because of shortages and high prices just after that war, the Japanese government's response was not to reduce protection but instead to extend it to its colonies and to shift from a national to an imperial rice self-sufficiency policy. That involved accelerated investments in agricultural development in the colonies of Korea and Taiwan behind an ever-higher external tariff wall that by the latter 1930s provided a nominal rate of

rice protection for the empire of more than 60 percent. After postwar reconstruction, Japan continued to increase its agricultural protection, just as countries in Western Europe did, but to even higher levels.

Meanwhile, in the 1950s a liberated South Korea and Taiwan adopted an import-substituting industrialization strategy that harmed agriculture. That strategy was replaced in the early 1960s with a more neutral trade policy that resulted in very rapid export-oriented industrialization in those densely populated economies. That export-led development strategy in turn imposed competitive pressure on the farm sector, which, just as in Japan in earlier decades, prompted farmers to lobby (successfully, as it happened) for ever-higher levels of protection from import protection in those newly industrialized economies as well (Anderson and Hayami 1986).

Policy Developments since the 1950s The historical tendency for countries to change gradually from taxing to subsidizing agriculture relative to other sectors in the course of their economic development has not been universal. The exceptions are rich countries with an extreme comparative advantage in agriculture (such as Australia and New Zealand) and poor countries with an extreme comparative disadvantage in agriculture (such as South Korea, as with Japan earlier, and some oil-rich states in the Middle East). Poor-country farmers also were disadvantaged by an antirural bias in public investments in infrastructure and human capital (education, health, agricultural research and development), and sometimes also by having to effectively finance urban consumer food subsidy programs (Pinstrup-Andersen 1988). Within the agricultural sector of each country, import-competing industries tended to enjoy more government support than those that were more competitive internationally (Krueger, Schiff, and Valdes 1988; Herrmann et al. 1992). The study by Krueger, Schiff, and Valdes also reveals that, at least up to the mid-1980s, direct disincentives for farmers such as agricultural export taxes were less important than indirect disincentives in the form of import protection for the manufacturing sector or

overvalued exchange rates, both of which attracted resources away from agricultural industries producing tradable products.

This pattern of distortions to incentives is wasteful from a global viewpoint, and detrimental to the vast majority of the world's poorest people, who are small farmers in developing countries. Currently fewer than 15 million relatively wealthy farmers in developed countries, with an average of 78 hectares per worker, benefit at the expense of not only consumers and taxpayers in those rich countries but also the majority of the 1.3 billion relatively impoverished farmers and their families in developing countries, who have to earn a living from just 2.5 hectares per worker, on average. The evolution from taxing to subsidizing farmers as countries develop suggests that, if left unchecked, agricultural protectionism would continue to spread to newly industrializing countries in the decades ahead as governments sought to protect domestic producers from import competition as the farm sector came under pressure to shrink in relative terms and, eventually, in terms of absolute numbers of people engaged.

Such a prospect contributed to the resolve of several groups in the 1980s to try to counter those political forces. For example, agricultural-exporting countries formed the Cairns Group and succeeded in ensuring that GATT members included an agreement on agricultural trade and subsidy reform in the Uruguay Round outcome. Over the same period, international financial institutions made a more concerted effort to encourage developing countries to reduce unilaterally their distortions against agriculture. Accession to preferential trading agreements and to the WTO, and the demise of communism, have helped in some cases too.

Several groups have attempted to measure the extent of distortions to agricultural incentives since the 1950s. Anderson and Hayami (1986) report annual nominal rates of protection estimates for East Asian and high-income countries from 1955 to the early 1980s. Krueger, Schiff, and Valdes (1988, 1991) calculate both direct and indirect distortions to agricultural incentives for 17 developing countries from the early 1960s to the mid-1980s.

Tyers and Anderson (1992) estimate that the global welfare cost of agricultural protection doubled during the 1980s. And the Organisation for Economic Co-operation and Development (OECD 2006) has provided annual estimates in a consistent way since 1986, for developed countries plus Korea, Mexico, and some of Europe's transition economies, of what are now called producer support estimates (PSEs) and consumer subsidy equivalents (CSEs) for agricultural and food products. These efforts use somewhat different methodologies, so there was a clear need to provide a comparable set of estimates for both developed and developing countries using a standard methodology. A World Bank research project took on that task, covering the period since the late 1950s for as many as 73 large economies. Together those countries account for about 90 percent of both global income and farm output. Below the methodology is briefly described and the findings summarized, details of which can be found in Anderson (2008).

Methodology for Estimating Distortions to Incentives The simplest indicator of distortions to agricultural incentives is provided by the nominal rate of assistance (NRA), defined as the percentage by which a country's government policies have raised (or lowered if it is negative) gross returns to producers above (below) what they would be without the government's intervention. The NRAs are based on estimates of assistance to individual farm industries. Most distortions to industries producing internationally tradable products come from trade measures, such as a tariff imposed on the import price or a tax or subsidy imposed on exports at the country's border. An ad valorem tariff or export subsidy is the equivalent of a production subsidy and a consumption tax expressed as a percentage of the border price, and that is what is captured in the NRA (and can also be captured in a consumer tax equivalent, or CTE) at the point in the value chain at which the product is traded. To get the NRA for the farmer, it is necessary to determine the extent of pass-through back along the value chain to the farm gate (and likewise forward to the consumer at the retail level to get the CTE). The NRA thus differs from the OECD's PSE, since

the PSE is expressed as a percentage of the distorted price and hence is lower than the NRA, which is expressed as a percentage of the undistorted price. But like the PSE, this NRA also captures any significant product-specific input price distortions by estimating their equivalence in terms of a higher output price and including that in the NRA for individual agricultural industries, and also adds non-product-specific distortions into the estimate for the overall sectoral NRA.

The degree of coverage of products for which NRA estimates are generated was 70 percent of the gross value of agricultural production (the same as for the OECD's PSE coverage). The OECD assumes the NRA for the residual noncovered products is the same as the average for covered products. However, in developing countries at least, policies affecting the noncovered products are often very different from those for covered products. The nontradables among them, for example, are often low-quality food staples that are subject to no distortionary policies. The World Bank project therefore provides three sets of estimates of the NRAs for noncovered farm products, one each for the import-competing, exportable, and nontradable subsectors. A weighted average for all agricultural products is then generated, using the values of production at unassisted prices as weights.

Since the late 1980s there has been a tendency in some high-income countries to move away from trade measures to more direct forms of assistance to farmers. This is largely in response to domestic pressures to reduce trade distortions to improve policy efficiency, as well as to pressures from abroad during and following the GATT's multilateral trade negotiations that resulted in the Uruguay Round Agreement on Agriculture. Some of those new measures are less decoupled from production incentives than others, so their production and trade effects have not entirely disappeared. And they still bestow a transfer on farmers. Hence it is helpful to show the NRA both with and without "decoupled" measures for those high-income countries adopting them.

It is not sufficient to look just at agricultural NRAs in isolation, because *relative* prices and hence relative rates of government assistance are what affect producers' incentives. In a model of a two-sector economy, an import tax has the same effect on the export sector as an export tax, and this carries over to a model that also includes a third sector producing only nontradables, to a model with imperfect competition, and regardless of the economy's size (Vousden 1990). The reason the result carries over with nontradables is that if an import tax of rate t is replaced by an export tax at rate t, all traded goods prices are reduced by $1/(1 + t)$ and therefore the price of all nontradables has to also change by that same amount if the market for nontradables (in which the quantity supplied domestically has to equal the quantity demanded) is to remain in equilibrium. Thus if one can assume there are no distortions in the markets for nontradables, the overall distortion to agricultural incentives can be captured by the extent to which the tradable parts of agricultural production are assisted or taxed relative to producers of other tradables. By generating estimates of the average NRA for nonagricultural tradables, it is then possible to calculate a relative rate of assistance, RRA, defined as:

$$RRA = 100[(1 + NRAag^t/100)/ (1 + NRAnonag^t/100) - 1]$$

where $NRAag^t$ and $NRAnonag^t$ are the average percentage NRAs for the tradable parts of the agricultural and nonagricultural sectors, respectively. Since an NRA cannot be less than -100 percent if producers are to earn anything, neither can the RRA. This measure is thus useful for providing an internationally comparable indication over time of the extent to which a country's policy regime has an anti- or proagricultural bias.

Caution is needed in interpreting the RRA. If it is negative for developing countries, and especially if it is positive for high-income countries, that probably indicates that farmers in developing countries effectively faced prices that were less than they would have been if there were no government distortions to goods markets. But there is the possibility that if all such distortions were removed globally   and even more so if they were removed only by developing countries   the international terms of trade could

deteriorate for developing countries. Whether that would be so severe as to make developing-country farmers worse off than without such reform is a moot point. One would need a series of global economy wide AGE models calibrated to each year in the time series to estimate the full general equilibrium effects on net farm incomes over time.

New Global Estimates of Distortions, 1955–2004 The World Bank research project's sample of 73 large economies comprises 22 high-income developed countries (HICs); 13 East European and Central Asian (ECA) economies in transition from socialism; and 17 African, 13 Asian, and 8 Latin American developing countries. Since these economies together account for around 90 percent of global income and of farm output, and since there are a similar number of developing countries as higher-income (including transition) countries, their NRAs and RRAs are very representative of each of those two groups and for the world as a whole.

Figure 1 illustrates the RRA estimates for the highest level of aggregation. It shows the strong upward trend in assistance to farmers in high-income countries from the mid-1950s to the late 1980s, rising from around 20 percent to 50 percent. Their rates have since come down somewhat, although less so when so-called decoupled assistance is included (the dashed uppermost line).

For developing countries there has also been a strong upward trend in the RRA estimate, but starting from a very low level. According to these estimates, the RRA averaged −50 percent in the latter 1950s, but this has risen steadily each half-decade since then. By the end of the 1990s the discrimination against agriculture by developing-country governments had virtually disappeared on average, and the RRA had even become slightly positive for the first time by the first half-decade of this century.

The weighted average for the two groups combined provides a global average RRA, shown as the middle line in figure 1, again with the so-called decoupled assistance included in the dashed upper part of that line. That line suggests global agricultural production was discouraged up to the mid-1980s, but
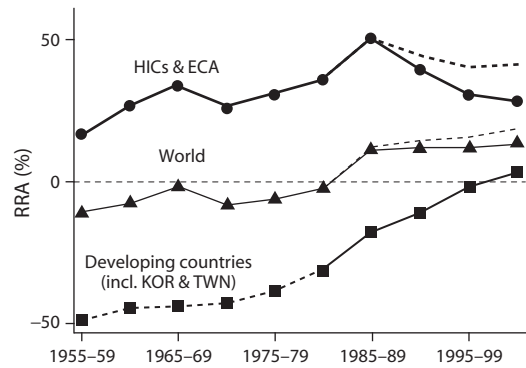
Figure 1

Relative rates of assistance to agricultural industries, showing weighted averages for high income countries, developing countries, and the world, 1955 to 2004 (percent). Source: Anderson (2008).

has since been encouraged on average: even though in the past two decades agricultural protection has come down a little in high-income countries, agricultural taxation in developing countries has been replaced by positive assistance that has been more than offsetting. There thus appears to be a convergence toward a positive RRA in both sets of countries.

These weighted averages of RRAs of course hide a great deal of diversity among countries. Within the high-income group the estimated RRAs in 2000 2004 range from virtually zero in Australia and New Zealand to more than 10 percent in North America, more than 30 percent in the European Union, and more than 100 percent in both the rest of Western Europe and Japan (and in Korea and Taiwan). For developing countries, the RRAs in 2000 2004 averaged 5 percent in Africa, 0 percent in Latin America, 1 percent in China, 6 percent in South Asia, and 20 percent in Southeast Asia. With even more diversity within each of those developing country regions, these estimates suggest there are many countries that could boost their overall economies by reducing those RRAs.

Within most countries including those with RRAs close to zero there is also still much diversity in rates of assistance to different farm industries. Almost all countries assist their import-competing farmers more than their more competi-

tive agricultural export industries. Given that exporters tend to be more productive and more innovative than other producers, this antitrade bias is slowing agricultural and overall economic growth.

As for the distribution across commodities, there is very strong support for the "rice pudding" products: rice, sugar, and milk. In high-income countries, rice has the dubious honor of being the most assisted, thanks especially to Japan. In 2000 2004 its NRA averaged 375 percent, compared with 125 percent for sugar and 65 percent for milk (with cotton and beef trailing close behind with NRAs of 55 and 45 percent, respectively). In developing countries the NRA for rice averaged "only" 17 percent in those years, but sugar and milk were higher at 41 and 32 percent, respectively. By contrast, developing countries continue to tax major export crops: their weighted average NRAs for cocoa, groundnuts, and cotton are still below zero.

Evidently there has been a great deal of reform in recent decades. The massive and widespread discrimination against farmers in developing countries that was identified empirically two decades ago in the seminal study by Krueger, Schiff, and Valdes (1988, 1991) has been much reduced and has even disappeared for some developing countries. Also, the growth in agricultural protection from the 1950s to the 1980s in high-income countries, identified empirically in Anderson and Hayami (1986), has slowed and, for the European Union, even reversed. There remains, however, plenty of scope for further welfare-enhancing reforms to policies distorting agricultural incentives. Reducing the remaining variance among rates of assistance to the various industries within each country's agricultural sector, and especially reducing the antitrade bias in current policies, would help. That in turn would reduce the wide range of NRAs across commodities globally, bringing down the present extreme positive rates for items such as rice, sugar, and milk and reducing the effective taxation of tropical cash crops such as cocoa.

The biggest uncertainty now is whether the RRAs for developing and high-income countries will continue their recent convergence and, if so, whether that will involve more reform by high-income countries

or a continuing upward long-run trend in RRAs for developing countries. Some commentators cite the tariffication and tariff bindings in the Uruguay Round Agreement on Agriculture as a reason to expect countries not to raise their agricultural assistance in the future. However, there is a great deal of "binding overhang" in those WTO commitments for many members and especially for developing countries (WTO, ITC, and UNCTAD 2007), meaning they could raise their applied import tariffs substantially on farm products before reaching the legally bound rates. It thus remains to be seen whether the forces of economic rationalism will be more or less powerful than the perennial agricultural protectionist forces in the decades ahead.

See also Agreement on Agriculture; agricultural trade negotiations; agriculture; Common Agricultural Policy; political economy of trade policy

FURTHER READING

Anderson, K., ed. 2008. *Distortions to Agricultural In centives: Global Perspective*. London: Palgrave Macmil lan. A synopsis of a new set of 75 country case studies providing annual estimates of agricultural protection and taxation over the past half century.

Anderson, K., and Y. Hayami, eds. 1986. *The Political Economy of Agricultural Protection: East Asia in Interna tional Perspective*. London: Allen and Unwin. A set of empirical studies on the extent and causes of agricul tural protection growth from the mid 1950s to the early 1980s in East Asia and other high income coun tries.

Anderson, K., W. Martin, and D. van der Mensbrugghe. 2006. "Distortions to World Trade: Impacts on Agricultural Markets and Farm Incomes." *Review of Agricultural Economics* 28 (2): 168 94. A global econo my wide computable general equilibrium modeling analysis, using the World Bank's Linkage model, of the effects of agricultural subsidies and import tariffs as of 2005.

Anderson, K., and E. Valenzuela. 2007. "Do Global Trade Distortions Still Harm Developing Country Farmers?" *Review of World Economics* 143 (1): 108 39. Another global economy wide computable general equilibrium modeling analysis, but using Purdue University's GTAP

model, of the effects of agricultural subsidies and import tariffs as of 2005.

Gulbrandsen, O., and A. Lindbeck. 1973. *The Economics of the Agricultural Sector.* Stockholm: Almquist and Wicksell. An early empirical study of trends in agricultural protection in Europe from the 1930s to the 1960s.

Haberler, G. 1958. *Trends in International Trade: A Report by a Panel of Experts.* Geneva: GATT. An early analysis of the problems the GATT contracting parties would face if, as was feared, agricultural protection continued to grow as manufacturing protection was being phased down in successive multilateral trade rounds.

Hermann, R., P. Schenck, R. Thiele, and M. Wiebelt. 1992. *Discrimination against Agriculture in Developing Countries?* Tubingen: J.C.B. Mohr. An early assessment of agricultural pricing policy in developing countries.

Kindleberger, C. P. 1975. "The Rise of Free Trade in Western Europe, 1820 1875." *Journal of Economic History* 35 (1): 20 55. A seminal study of the differences across Europe in the evolution of agricultural protection in the 19th century.

Krueger, A. O., M. Schiff, and A. Valdes. 1988. "Agricultural Incentives in Developing Countries: Measuring the Effect of Sectoral and Economywide Policies." *World Bank Economic Review* 2 (3): 255 72. A synopsis of the extent of taxation of agriculture in developing countries reported more fully in Krueger, Schiff, and Valdes (1991).

———. 1991. *The Political Economy of Agricultural Pricing Policy.* Vol. 1, *Latin America*; Vol. 2, *Asia*; and Vol. 3, *Africa and the Mediterranean.* Baltimore: Johns Hopkins University Press for the World Bank. A milestone empirical study of 18 developing countries, analyzing the evolution of direct and indirect distortions to agricultural incentives from the early 1960s to the mid 1980s and the political economy forces behind them.

Lindert, P. 1991. "Historical Patterns of Agricultural Protection." In *Agriculture and the State*, edited by P. Timmer. Ithaca: Cornell University Press. The best summary available of the very long history of agricultural protectionism in the current high income countries.

OECD. 2006. *Producer and Consumer Support Estimates, OECD Database 1986 2004.* Available at http://www.oecd.org/document/54/0,2340,en 2649 33727 350 09718 1 1 1 1,00.html. The definitive annual report of PSEs and CSEs in all the OECD member countries during the past two decades, but treating the European Union as one entity.

Pinstrup Andersen, P., ed. 1988. *Food Subsidies in Developing Countries: Cost, Benefits, and Policy Options.* Baltimore: Johns Hopkins University Press. A useful analysis of the consumer side of distortions to agricultural incentives in developing countries, complementing the majority of other studies that focus mainly on the producer impacts of agricultural policies.

Tyers, R., and K. Anderson. 1992. *Disarray in World Food Markets: A Quantitative Assessment.* Cambridge: Cambridge University Press. A model based analysis of distortions to markets for grains, sugar, and livestock products, a preliminary version of which was used as the basis of the World Bank's *World Development Report 1986.*

Vousden, N. 1990. *The Economic Theory of Protection.* Cambridge: Cambridge University Press. A seminal textbook on the theory of protectionist international trade policies.

WTO, ITC, and UNCTAD. 2007. *World Tariff Profiles 2006.* Geneva: WTO, ITC, and UNCTAD. A new joint publication by these three international institutions that makes publicly available, in an easily accessible form, the WTO legally bound and the applied tariff structures of more than 100 countries.

KYM ANDERSON

■ Doha Round

The Doha Round, launched in 2001 and formally named the Doha Development Agenda (DDA), was the first round of global trade negotiations held under the auspices of the World Trade Organization (WTO). To understand the dynamics of the Doha Round, it is important to understand its genesis.

The creation of the WTO was one of the major results of the Uruguay Round of multilateral negotiations. The WTO began operations on January 1, 1995. Its early years were dominated by negotiations to finalize agreement on several services sectors (finance, telecommunications, maritime transportation) and to liberalize trade in information technology (IT) products. With the exception of the

maritime services talks, these were successful: telecom and financial services agreements were concluded in 1996; the Information Technology Agreement (ITA) a so-called zero-for-zero deal under which signatories abolished tariffs on a common set of IT products entered into force in 1997.

Flush with these achievements, the first ministerial conference of the WTO, held in Singapore in December 1996, agreed to a work program for the following years. Members of the Organisation for Economic Co-operation and Development (OECD) proposed to put government procurement, trade facilitation, competition law, and policies on foreign investment on the WTO agenda. A minority (led by France and the United States) also sought to launch discussions on labor standards. Opposition by many developing countries kept labor standards off the WTO agenda, but members did agree to create working groups to discuss and study the relationship between trade and competition and investment policy disciplines, transparency in government procurement, and trade facilitation. These subsequently came to be known as the "Singapore issues."

In the years that followed, OECD members failed to agree on a Multilateral Agreement on Investment; furthermore the East Asian financial crisis and follow-on crises in other parts of the world added to skepticism on the part of many developing country governments and numerous vocal nongovernmental organizations that the WTO was a balanced and "fair" organization.

Research supporting aspects of this skepticism began to emerge in the late 1990s, concluding that the net benefits for many developing countries in narrow market-access terms were limited and may even have been negative as the result of the new requirements under the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) to enforce patent rights and other forms of protection for intangible assets. TRIPS was the first example of a multilateral trade agreement that involved a significant element of policy harmonization. Although most economists accept that there is an economic logic to protection of intellectual property, both theory and economic history suggest that

harmonization is unlikely to be an optimal outcome for all countries, in particular poor economies. Whatever one's view of the economics, TRIPS greatly increased the awareness of many developing countries of the need to carefully scrutinize the likely impacts of agreements that entail regulatory harmonization.

A major feature of the WTO was that developing countries became subject to numerous rules of the game that they had not been affected by as members of the WTO's predecessor, the General Agreement on Tariffs and Trade. The costs of implementation of some of these WTO disciplines, if defined not just narrowly in terms of required legal and regulatory changes, but in terms of what is necessary to benefit from them, could be significant. Assistance to meet the costs of implementation was a matter for governments to request from national and international development agencies. Many developing countries argued that the provision of such assistance had been inadequate.

These considerations help explain why only a few years after the establishment of the WTO, efforts were made to initiate a new round of trade negotiations. The requirement in the Uruguay Round agreements on agriculture and services to initiate new negotiations in these two areas in 2000 provided an additional rationale for launching a broader round, as this would allow for cross-issue linkages and trade-offs.

Countries had very different objectives, with some (mostly developed countries) seeking to build on the Singapore ministerial and launch talks on the new areas as well as the core market-access agenda, and others (mostly developing nations) seeking to address implementation and related problems associated with existing agreements.

The attempt to launch a "Millennium Round" at the 1999 WTO ministerial meeting in Seattle failed to achieve the necessary consensus, largely reflecting opposition by many developing countries to engage in new negotiations when they had yet to digest the Uruguay Round and a perception that implementation concerns were not being taken sufficiently seriously, as well as resistance on the part of the

European Union (EU) and several other OECD countries to engaging in talks to significantly reduce support for their farmers. Many developing countries also opposed proposals to negotiate WTO disciplines for the Singapore issues, environmental policies, and (minimum) labor standards. Notwithstanding the deep differences in views, additional preparatory discussions in 2000 2001 allowed the first round of multilateral negotiations under WTO auspices to be launched at the 2001 Ministerial Conference in Doha, Qatar.

Major Players and Issues The name given to the round the Doha Development Agenda arose for two reasons. First, many WTO members held the view that a multilateral effort to reduce trade-distorting policies was part of an appropriate response to the terrorist attack on New York and Washington, DC, on September 11, 2001. The view that trade cooperation could help address some of the causes of terrorism by promoting economic growth was strongly espoused by the then U.S. trade representative Robert Zoellick and helped to change the political atmospherics that up to that point had not generated sufficient support for the launch of new trade talks. Second, the name reflected recognition of the view of many developing countries that the Uruguay Round had been "unbalanced." Making "development" a prominent goal of the round had a major impact on the subsequent negotiating dynamics.

Reflecting the differences in objectives of the participants, the negotiating agenda of the round was multidimensional. It spanned market access for goods and services, the WTO's existing rules, potential new disciplines on the Singapore issues, implementation problems, and a renewed emphasis on "special and differential treatment" (SDT) for developing countries.

The major protagonists were the EU and the United States on the side of OECD countries, and Brazil and India on the side of the developing countries. In addition to these four giants, a variety of coalitions played an important role, both in the run-up to the Doha Round and during the negotiations. Influential groups included the so-called Like Minded Group of developing countries, which argued for more SDT, for addressing implementation concerns, and for nonreciprocity in negotiations; the Africa Group, which espoused a similar position; the least-developed countries (LDC group) and a variety of "G-x" groups, where "x" is a number indicating how many countries were participants in the launch of the group. The most prominent of these was the G20, a group of countries led by Brazil and India that also included China and South Africa. The G20 had a fluctuating membership, with some countries leaving sometimes as the result of pressure (inducements) offered by the United States (the departure of several Central American countries following the formation of the Central American Free Trade Agreement being an example) and others joining. Other groups included the G90 comprising almost the totality of developing countries in the WTO and the G11, a group of developing countries that were active in the nonagricultural market-access talks.

As many developing countries were not convinced that launching negotiations on the Singapore issues was in their interest, a last-minute compromise reached at the Doha Ministerial Conference was that negotiations on these four subject areas would commence at the 2003 meeting of WTO ministers. Based on an intervention by India, it was specified that a precondition for such negotiations was an "explicit consensus" among WTO members on their modalities.

The EU, Japan, and the Republic of Korea were the primary *demandeurs* for negotiations on all four Singapore issues. Three groups of developing countries the African Union; the LDCs; and the African, Caribbean, and Pacific group of countries had all agreed at the ministerial level in 2003 that they did not support launching negotiations on any of these topics. They were joined by a number of middle-income countries, such as Malaysia. Some countries argued that these were marginal issues, with only limited benefits for developing countries; that they could give rise to potentially significant implementation costs; and that they would divert scarce negotiating resources and political attention away from the more important

market-access agenda. Although many middle-income economies, including most Latin American countries, did not have serious concerns about launching negotiations on the four subjects, it became clear in 2003 that an "explicit consensus" did not exist. As a result, the 2003 Cancún Ministerial Conference failed   the second one to do so in the WTO's short history.

The failure was not due solely to differences on the Singapore issues. Disagreements were particularly prominent on agriculture. Developed countries' farm subsidies   with the EU's Common Agricultural Policy (CAP) and the U.S. system of farm subsidies being the most prominent   were major bones of contention. Opposition by the EU to liberalization of its agricultural policies and stronger disciplines on farm subsidies was a fundamental factor impeding agreement. In addition, U.S. unwillingness to address demands by West African cotton producers to reduce production subsidies as a priority also played an important role in souring the atmosphere in Cancún.

A "Framework Agreement" negotiated in July 2004 removed three of the Singapore issues from the negotiating table (retaining only trade facilitation), specified that LDCs were not expected to make any market-access concessions (taking up then EU trade commissioner Pascal Lamy's call that these countries should benefit from the round "for free"), and established negotiating frameworks for the key areas of agriculture and nonagricultural market access. It was agreed that formulas were to be used to reduce trade barriers in both areas, and that export subsidies in agriculture were to be prohibited by a specific date to be negotiated.

One result of the decision to take three of the Singapore issues off the table was that to a much greater extent than in the Uruguay Round negotiations centered on tariffs. Although in principle services were another important market-access negotiation, WTO members were not willing to make significant liberalization commitments   in part because of a desire first to determine what the contours of a possible deal would be for merchandise trade, agriculture in particular. A major challenge

confronting negotiators was to agree to substantially lower remaining tariff peaks. As average tariff rates in most industrial countries are now relatively low, the main payoff from both a development and an economic efficiency perspective would come from reducing the dispersion in tariff protection by lowering the highest tariffs more than the average. A straightforward way of doing this is to apply a nonlinear tariff reduction formula to each country's prevailing tariffs.

WTO members agreed to make tariff-cutting formulas a core negotiating modality in the Doha Round, with eventual emergence of variants of the so-called Swiss formula   first used in the Tokyo Round (1973 79)   as the basis of discussion for merchandise trade liberalization. The Swiss formula is defined as $T_1 = MT_0/(M + T_0)$, where T_1 is the new tariff commitment, T_0 is the prevailing one, and M is a coefficient   the maximum level of the permitted tariff. The formula generates nonlinear cuts, with higher proportional reductions for higher tariffs. It was decided that (at least) two coefficients would be used, one for developed countries and another for developing nations, such that developed countries would reduce their tariffs proportionally more in accordance with the notion of SDT for developing countries. Limited exemptions would be permitted to allow governments to maintain higher levels of protection for a subset of "special" and "sensitive" products. Given that the major developing country coalitions such as the G20 spanned a differentiated set of countries, insistence on undifferentiated SDT for all members made it more difficult to agree on a coefficient and resulted in much attention also being given to the criteria that would apply for possible exemptions from the formula for "sensitive" products.

Disagreements on the specification of the formulas and the magnitude of allowable exceptions for specific products could not be overcome during 2004 7. Negotiations between the so-called five interested parties in agriculture Australia and Brazil, representing export interests, the EU and India, representing the "defensive" interests, and the United States (both an exporter and a provider of

substantial support to its farmers) did not manage to identify a package of mutually acceptable reforms. Much of the discussion centered on technical issues such as how to go about converting specific tariffs (taxes that are based on volume rather than the value of imported products) into ad valorem equivalents. Although these were important, the fact that such technical matters needed to be addressed by ministers illustrated the depth of the political sensitivities that prevented lower-level officials from agreeing on negotiating modalities in Geneva.

The 2005 ministerial meeting in Hong Kong did not lead to any major breakthroughs, with the exception of (conditional) agreement to ban export subsidies for agricultural products in 2013, promises of an increase in development assistance ("aid for trade"), and a promise by rich countries to extend duty-free, quota-free access to their markets for at least 97 percent of exports originating in LDCs. As noted by many commentators, the last promise did little to go beyond the status quo, and the exclusion of 3 percent of tariff lines was enough to permit countries to maintain tariffs on the most competitive LDC exports if they desired to. Ministers renewed their commitment to conclude the DDA by the end of 2006. A roadmap for key decisions was agreed on, including a target date of April 30 to agree on modalities (formulas and liberalization parameters) for agriculture and nonagricultural merchandise trade. Comprehensive draft schedules based on these modalities were expected to be submitted by July 31, 2006. This was also the deadline for a second round of revised market opening offers in services.

As of mid-2006, the contours of a possible agreement on merchandise trade liberalization had emerged. It was described as a 20-20-20 package by the director general of the WTO. Under this proposal the EU would cut its bound agricultural tariffs by an estimated 54 percent (following the proposal made by the G20), the United States would cap its total trade-distorting agricultural subsidies at a maximum of $20 billion, and developing countries would agree to a maximum tariff on manufactures of 20 percent. A deal along the lines of the 20-20-20 package would have implied a significant reduction not only in tariff bindings but also in applied policies. Negotiators could not agree on a specific compromise, however. The G20 sought deeper cuts in EU tariffs and both the G20 and the EU wanted the United States to accept a lower ceiling on its domestic farm subsidies.

By this time the original endpoint envisaged for the Doha Round (January 1, 2005) had been missed by a wide margin, as were nearly all deadlines that had been set during the talks. This pattern extended into 2006 7. The inability of the major protagonists to make concessions led the director general of the WTO to suspend the negotiations in mid-2006. Despite efforts in early 2007 to resuscitate the process, which led to a high-profile meeting between four of the major protagonists (the EU, the United States, Brazil, and India) in Potsdam in June, no agreement could be reached on how much to cut farm subsidies and tariffs on agricultural and industrial goods. Developing countries remained disappointed by the United States' offer on agricultural subsidies and the EU's unwillingness to open its agricultural markets. Conversely, the EU and the United States were unsatisfied by what they regarded as inadequate offers by Brazil and India on industrial tariffs. At the time of writing (April 2008), prospects for an imminent conclusion of the Doha Round appeared dim. The expiry of Trade Promotion Authority a provision previously called "Fast Track," which precludes the U.S. Congress from introducing amendments to a multilaterally negotiated deal in June 2007 removed an important focal point for a timely conclusion of the talks. The political calendars of several major players suggested that the Doha Round was unlikely to conclude before 2009 10.

Progress and Prospects Despite the torturous negotiating process, some progress was made during the Doha Round to deal with several development-related concerns that had become prominent in the period immediately following the creation of the WTO. Thus in a Declaration on the TRIPS Agreement and Public Health, developing countries without the capacity to produce pharmaceuticals

were permitted to import generics from countries that do have capacity under compulsory licensing arrangements.

A consensus also emerged around the notion that trade negotiations should be complemented by assistance for developing countries, both to help address implementation costs associated with specific WTO disciplines and, more generally, to bolster the competitiveness of domestic firms through actions to reduce the costs of trade and doing business. The recognition that trade negotiations must be complemented by actions to help address the prevailing constraints in many poor countries that inhibit exploitation of trade opportunities was an important development. Insofar as the Doha Round has led development organizations to focus greater attention on the trade agenda and constraints in developing countries, this should be counted as a positive outcome.

Why so little success in obtaining agreement? A number of reasons can be identified. Many of these revolve around the argument that there was not enough on the table to mobilize support within the OECD countries to accept greater agricultural liberalization. Concerns on the part of a number of smaller developing countries that they would suffer losses from the erosion of trade preferences following agreement among the larger WTO members to reduce tariffs on a most-favored-nation basis also played a role.

It is important to consider that the Doha Round took place during a period in which there was a boom in world trade, in part driven by trade liberalization that was implemented unilaterally by governments. World Bank research suggests that the uniform tariff equivalent of applied tariffs for merchandise products in 2007 was 7 percent in high-income countries and 15 percent in developing countries. The numbers are heavily influenced by the substantially higher protection of agricultural products, around 40 and 25 percent, respectively. They are much lower for manufactures (4 and 14 percent).

The implication of the relatively low levels of applied tariffs in many countries is that there is less

to play for. This translates into fewer export interests having an incentive to invest significant (political) resources in engaging in trade talks and providing the political support that is needed for liberalization. As most protection is now concentrated in agriculture, matters are complicated further by the fact that this sector is of significant export interest to only a subset of WTO members. Less than 10 percent of world trade is in agricultural products. Although trade volumes no doubt would be higher if rates of protection were lowered, it is trade in other merchandise that dominates, and much of this trade is relatively free.

Because WTO negotiations are barter exchanges, concessions in agriculture by high-income countries need to be balanced by concessions on the part of the agricultural exporting nations that will benefit in mercantilist, export volume terms. Although average levels of protection have fallen in these countries, there are still significant barriers to trade. The low prevailing averages mask relatively high tariff peaks in both agriculture and manufactures in many countries. In services, despite significant liberalization of trade and investment in recent years, numerous barriers persist. Finally, it is important to consider that the negotiating coin of the WTO is policy bindings the levels of protection that a country commits itself not to exceed. These bindings are valuable even if a specific commitment does not imply much, if any, reduction in applied levels of protection, as they create greater certainty that past liberalization will not be easily reversed.

In principle it would appear therefore that even with the relatively narrow market-access (tariff) agenda that dominated the Doha Round after 2004, a bargain should have been feasible, with "payment" for OECD agricultural liberalization taking the form of traditional market-access liberalization by major developing countries. In practice, however, developing countries appeared unwilling to offer enough in terms of liberalization and bindings to induce the OECD countries to move enough to allow a deal to be struck. This in turn suggests that the major players perceived their best alternatives to a negotiated agreement to dominate an outcome

under which they would have had to make additional concessions in order to get more liberalization from their major trading partners.

What the outcome of the Doha Round will be remains to be seen. The same is true regarding the consequences of the Doha Round for the WTO as an institution. The difficulty in attaining a consensus to negotiate on the Singapore issues, and the limited progress that was made on extending the coverage of services commitments, suggest there may be limits regarding the set of policies that can (should) be subject to binding multilateral disciplines.

See also agricultural trade negotiations; multilateral trade negotiations; special and differential treatment; Uruguay Round; World Trade Organization

FURTHER READING

Anderson, Kym, and Will Martin, eds. 2006. *Agricultural Trade Reform and the Doha Development Agenda*. Basingstoke, UK: Palgrave Macmillan. Analyzes the potential economic consequences of agricultural liber alization and the specific aspects of the Doha talks on agriculture.

Hertel, Thomas, and L. Alan Winters, eds. 2006. *Poverty and the WTO: Impacts of the Doha Development Agenda*. Basingstoke, UK: Palgrave Macmillan. Focuses on the potential impacts of global trade reforms on poverty within countries.

Hoekman, Bernard, Aaditya Mattoo, and Philip English, eds. 2002. *Development, Trade, and the WTO: A Handbook*. Washington DC: World Bank. A compila tion of short papers that describe the WTO and the subjects that were on the Doha negotiating agenda.

Lee, Donna, and Rorden Wilkinson, eds. 2007. *The WTO after Hong Kong: Progress in, and Prospects for, the Doha Development Agenda*. London: Routledge. Provides an international relations perspective on the negotiations, including discussions of coalitions and negotiating strategies.

UN, Millennium Project. 2005. "Trade for Develop ment." Task Force on Trade. New York: United Nations Development Program. Includes a comprehensive analysis of the issues on the Doha agenda and their im portance from an economic development perspective.

World Trade Organization (WTO). http://www.wto.org/. The WTO home page is a primary source of information on developments in the negotiations.

BERNARD HOEKMAN

■ dollar standard

In the absence of a purely international form of money, a strong national currency is central to international commerce. Outside Europe, the dollar dominates as a vehicle currency in interbank foreign exchange, as an invoice currency in international trade and capital flows, and in official exchange reserves. How did the dollar establish its preeminent international role?

After World War II, the United States had the world's largest economy and the only stable financial system with open foreign exchange markets. Inflation and exchange controls proliferated throughout Europe as well as in Japan and most developing countries. Thus the dollar naturally became the world's vehicle currency for (private) interbank transacting between any pair of national currencies and became the intervention currency that governments used for stabilizing their exchange rates. Foreign governments as well as private corporations could build up their dollar reserves knowing that they had a liquid market in which to sell them.

The Bretton Woods agreement of 1945 formally ratified the dollar as the dominant currency. Under Article IV, countries outside the communist bloc declared dollar parities, while, as center country, the United States had a passive foreign exchange rate policy except for its residual tie to gold.

The Dollar as Facilitator of International Exchange But why does the dollar standard continue even when most other industrial countries, such as Japan and those in Europe, no longer have exchange controls and Article IV was amended in 1976 so that member countries no longer need to declare dollar parities? Suppose there are 150 national currencies in the world economy. To facilitate international exchange, the markets themselves would always pick just one as the central money for

intermediating nearly all international payments. Transaction efficiency increases because the number of active bilateral foreign exchange markets necessary for clearing international payments, both spot and forward, is thereby reduced.

A little algebra helps explain continued dollar predominance in a world of N national currencies. The total number of country pairs in the system is the combination of N things taken two at a time ($^{N}C_2$). If foreign exchange dealers tried to trade across each pair — say, Swedish kronor against Australian dollars, or Korean won against Japanese yen — there would be a huge number of active foreign exchange markets. With 150 national currencies in the world (N = 150), if dealers tried to trade each pair bilaterally that would lead to 11,175 different markets!

It is expensive for any bank to set up a foreign exchange trading desk. Thus, rather than trading all pairs of currencies bilaterally, in practice just one currency, the Nth, is chosen as the central vehicle currency. Then all trading and exchange takes place first against the vehicle currency before going to the others. By having all currency trading against that one currency, the number of markets in the system can be reduced to N-1. Thus, with 150 countries, we need just 149 foreign exchange markets against the U.S. dollar (instead of 11,175 bilateral exchange rates against each other). Unlike the Bretton Woods system, where all countries other than the United States set official dollar parities, this result does not depend on any formal agreement among governments. In private markets today, choosing one currency such as the dollar to be the intermediary currency is the most natural way of economizing on foreign exchange transacting.

But history is important. If one country starts to provide the central money, as the United States did in the late 1940s, it becomes a natural monopoly because of the economies of scale. The more countries that deal in dollars, the cheaper it is for any one of them all to deal in dollars as unit transaction costs fall — particularly in forward exchange markets that are much thinner and potentially much less liquid than spot markets. If you are a Japanese importer of Swedish Volvos, to pay for the Volvos you would first have your bank convert your yen into dollars on the open market, then use the dollars to buy Swedish kronor. Volvo Corporation receives the Swedish kronor, and the importer gets the Volvos. But the dollar is the intermediary currency: two interbank transactions involving dollars are cheaper than trying to trade yen directly for Swedish kronor at different terms to maturity.

Using the standard textbook classification of the roles of money, table 1 summarizes our paradigm of the dollar's central role in facilitating international exchange. For both the private and government sectors, the dollar performs as medium of exchange, store of value, unit of account, and standard of deferred payment for international transactions on current and capital accounts — as it has done since 1945.

First, the dollar is a *medium of exchange.* Because the foreign exchange markets are mainly interbank, the dollar is the vehicle currency in interbank transactions serving customers in the private sector. The dollar is on one side or the other of about 90 percent of interbank transactions worldwide. Thus, when any government intervenes to influence its exchange rate, it also finds it cheaper and more convenient to use the dollar as the official intervention currency. (The major exception to this convention is a fringe of European countries to the east of the euro area, which mainly use the euro as their central money.) Although the dollar is the predominant money in foreign currency trading, London has the largest foreign exchange market — bigger than New York's. And large dollar-based markets also exist in Singapore and Hong Kong.

Table 1
The U.S. dollar's facilitating role as international money since 1945

	Private	Official
Medium of exchange	Vehicle	Intervention
Store of value	Banking	Reserves
Unit of account	Invoice	Peg
Standard of deferred payment	Private bonds	Sovereign bonds

Second, the dollar is an international *store of value.* Corporations and some individuals hold dollar bank accounts in London, Singapore, and other offshore banking centers—as well as in the United States itself. It is estimated that more than half the stock of coin and currency issued by the U.S. government circulates abroad in Latin America, Russia, Africa, and other financially distressed areas. In developing countries, as a whole, about two-thirds of their official exchange reserves are in dollars.

Third, the dollar serves as a *unit of account* for much of international trade. Trade in primary commodities shows a strong pattern of using the dollar as the main currency of *invoice*. Exports of homogeneous primary products such as oil, wheat, and copper all tend to be invoiced in dollars, with worldwide price formation in a centralized exchange. Spot trading, but particularly forward contracting, is concentrated at these centralized exchanges—which are usually in American cities such as Chicago and New York, although dollar-denominated commodity exchanges also exist in London and elsewhere.

Invoicing patterns for exports of manufactured goods are more complex. Major industrial countries with strong currencies tend to invoice their exports in their home currencies. Before the establishment of the European Monetary Union, more than 75 percent of German exports had been invoiced in deutsche marks, more than 50 percent of French exports invoiced in francs, and so on. (McKinnon 1979 provided a rationale for this difference in invoicing practices between primary commodities and manufactures.)

Within Asia, however, foreign trade is invoiced mainly in dollars. In Korea, for example, the proportion of imports invoiced in U.S. dollars is about 80 percent, and even higher for exports. Other less mature economies in East Asia, such as China, use dollars even more intensively in invoicing their foreign trade. In Japan, the most industrialized and richest economy in East Asia, almost half of its exports and three-quarters of its imports are invoiced in dollars.

Fourth, as *a standard of deferred payment*—which is also a traditional role of money—private

and sovereign bonds in international markets are heavily denominated in U.S. dollars, though the euro seems to be as important. The growth of a broadly based euro-denominated bond market within Europe has made it much more attractive for foreigners to borrow by issuing euro bonds. The euro area is unusual in that it is a net creditor in the world economy that can lend in its own currency. Other net creditors are more or less confined to lending in dollars.

Despite the increasing importance of the euro in international bond markets, U.S. Treasuries are still taken as the benchmark, or "risk-free" asset. That is, dollar-denominated sovereign bonds issued by emerging markets the world over have their credit ratings measured relative to U.S. Treasuries. Thus, risk premiums in interest rates on these bonds are typically quoted as so many percentage points over U.S. Treasuries.

The Dollar as Nominal Anchor Beyond the traditional roles of money outlined in table 1, the dollar has a second and complementary international function. Foreign monetary authorities may better anchor their own domestic price levels by choosing to peg, officially or unofficially, to the dollar. By opting to keep their dollar exchange rates stable, foreign governments are essentially opting to harmonize—without always succeeding—their monetary policies with that of the United States. This monetary harmonization has two avenues: (1) international commodity arbitrage—the *arbitrage avenue*, and (2) the *signaling avenue*, where other central banks take their cue from actions of the U.S. Federal Reserve Bank.

For the dollar to function successfully as nominal anchor, however, two important conditions must be satisfied: (1) The U.S. price level, as measured by a broad index of tradable goods prices, is stable and expected to remain so. (2) Most countries, and certainly neighboring ones, are on the same international standard, that is, they also fix their exchange rates to the dollar.

In the history of the post–World War II dollar standard, these two conditions were satisfied in some periods but not in others. Indeed, in contrast to the

dollar's continued robustness as the facilitator of international exchange under either fixed or floating exchange rates, its function as nominal anchor has continually metamorphosed. It was strong in the Bretton Woods period of the 1950s and 1960s of fixed dollar parities, weakened greatly in the inflationary period of the 1970s into the 1980s, and then strengthened in the 1990s into the new millennium when the American price level again became fairly stable. The current proclivity of most Asian economies (with the major exception of Japan) and many emerging markets elsewhere to peg to the dollar, if only softly and informally, is sometimes called Bretton Woods II.

Consequences of International Currency Asymmetry Developing countries and a few industrial ones cannot borrow internationally in their own currencies   a phenomenon that has been dubbed *original sin*. Those developing countries that are dollar debtors live on sufferance: their domestic monies are only "provisional." Any economic or political disturbance at home provokes the suspicion that these foreign currency debts may not be repayable and that the domestic currency will depreciate against the dollar   as in the East Asian crisis of 1997 98.

But what about the biggest international debtor of all? After running trade deficits for more than 25 years, the United States is a net debtor: its liabilities in 2006 exceeded its claims on the rest of the world by about $4 trillion. At about 30 percent of U.S. gross national product (GNP) in 2006, America's net international indebtedness was higher than that of any other industrialized country   and higher than, say, Brazil's, which was only 20 percent of Brazilian GNP. Yet, unlike Latin American currencies today, and unlike East Asian currencies in 1997 98, the dollar is not threatened by a loss of confidence. As long as its purchasing power is seen to be stable, that is, as long as the Federal Reserve keeps price inflation very low, the dollar cannot be attacked in the usual sense. Default risk associated with foreigners holding dollar assets is virtually nonexistent. If need be, the Fed could always print more money to redeem U.S. Treasury bonds outstanding. Thus the United States has an indefinitely long line of credit with the rest of the world.

But this virtual invulnerability of the center country as debtor to foreign exchange risk means that this risk shifts to creditor countries that, Europe aside, cannot *lend* to the United States in their own currencies   a circumstance known as *conflicted virtue*. Creditor countries such as China and Japan with trade surpluses, reflecting their high domestic savings, simply accumulate large volumes of liquid dollar claims on the United States, leaving them vulnerable to unwanted currency appreciations. Conflicted virtue in creditor countries is the mirror image of original sin in debtor economies. Both creditor and debtor economies are exposed to serious currency risks should their currencies fluctuate against the dollar. The only country exempted from exchange risk is the United States itself.

See also Bretton Woods system; conflicted virtue; dominant currency; euro; exchange rate regimes; gold standard, international; interest parity conditions; international reserves; original sin; reserve currency; twin deficits; vehicle currency

FURTHER READING

McKinnon, Ronald. 1979. *Money in International Exchange: The Convertible Currency System*. New York: Oxford University Press.

   . 2005. *Exchange Rates under the East Asian Dollar Standard: Living with Conflicted Virtue*. Cambridge, MA: MIT Press. Introduces and describes in detail the concept of conflicted virtue.

RONALD MCKINNON

■ domestic content requirements

Domestic content requirements compel firms to purchase a certain percentage of their inputs from domestic firms as a precondition for local market access or preferential policy treatment. In general, domestic content requirements act as a protectionist measure, since they usually improve the competitive position of domestic firms in relation to foreign firms. Nonetheless, the ultimate effect of domestic

content requirements depends on the form of the requirements, the characteristics of demand, market structure, and the nature of the production process.

Motives and Form In their simplest form domestic content requirements oblige firms to purchase a certain percentage of their inputs from domestic firms. If a firm fails to meet the requirement, it may be denied market access or its exports may be hit with additional tariffs. Alternatively, the firm may fail to qualify for policy benefits such as production subsidies. Although domestic content requirements usually apply to all firms, they typically have no impact on domestic firm production decisions, since domestic firms generally use a higher percentage of domestic inputs in their production than foreign firms do, and the political economy determinants of these measures generally result in domestic content requirements that surpass the natural domestic content selected by unconstrained foreign firms but are met or exceeded by the natural domestic content selected by domestic firms.

The implementation of domestic content or local content requirements is usually motivated by a country's desire to assist domestic firms or to increase the level of domestic economic activity. For example, many developing countries used domestic content regulations as an element of their import substituting industrialization policies. Domestic content regulations have also been imposed to influence the activities of foreign investors, with the goal of increasing foreign firm investment in the domestic economy or of increasing the domestic share of input purchases by foreign firms that already operate in the domestic market. Such policies have been especially common in large manufacturing industries, such as automobile assembly, where domestic countries preferred that investors provide many jobs in the local economy, rather than simply doing a small amount of local assembly using foreign materials. Since the imposition of local content requirements on foreign investors violates the national treatment principle of the General Agreement on Tariffs and Trade, however, the Uruguay Round included a trade-related investment measures agreement, which requires countries to remove domestic content re-

quirements on foreign investment. Nonetheless, such requirements may still remain in place, though in less transparent forms.

Content regulations have proliferated in a regional context as countries have entered into an ever-increasing number of free trade agreements. Since free trade agreements eliminate tariffs on all trade among member countries while leaving tariffs on nonmembers unchanged, free trade agreements create an incentive for tariff-shopping by nonmember exporters that wish to sell their products in the member market. To prevent tariff-shopping, free trade areas typically include rules of origin that stipulate the percentage of area content or area value added that the traded product must contain if it is to qualify for the preferential tariff treatment extended to member-country products.

Content preferences are generally implemented to protect domestic or regional markets. However, the Generalized System of Preferences (GSP), which provides tariff reductions for products exported from developing countries to developed countries, usually requires products to contain a minimum level of domestic content from the developing country. Finally, it is important to note that domestic content requirements are also imposed in service sectors, where regulations may require, for example, that a certain percentage of broadcast programming originates from domestic providers.

Impact of Content Requirements When domestic inputs are of similar quality and lower price than foreign inputs, firms will willingly meet or exceed the domestic content requirement and the requirement will have no effect on firm decisions. Since domestic content requirements are set in a political environment, however, they generally force foreign firms to purchase a larger fraction of their inputs from domestic sources than they would if they were unconstrained, while domestic content requirements usually have no effect on the sourcing choices of domestic firms. The best response for a cost-minimizing foreign firm is to exactly meet, though not exceed, the purchase criteria of the domestic content requirement. Alternatively, the foreign firm may decide to pay the tariff penalty that is associated with

noncompliance if the tariff cost is less than the extra cost of purchasing more domestic inputs. Either way, since domestic content requirements raise the relative production costs of foreign firms, they benefit domestic producers of final goods by increasing domestic firm sales or profits. In addition, they may reduce competition in the final goods market if they reduce the volume of goods sold by the foreign firm in the domestic market.

Whether they are imposed by the domestic country or by an importing developed country that is administering its GSP tariff preferences, domestic content requirements protect domestic intermediates producers. Since these policies increase production costs, domestic content regulations harm domestic customers, who face higher prices for the goods they buy. In addition, programs such as the GSP have an ambiguous effect on developing country welfare, as the increased demand for imports of the final good from the developing country will be offset, and possibly even reversed, by the cost increases that are associated with the content requirement.

Domestic content requirements take many different forms. For this reason, the form of the requirement has great influence on its economic effects. For example, domestic expenditure on capital is often excluded, while domestic labor is generally counted toward the domestic content requirement. When this is the case, firms respond to the relative cost incentive by investing in more labor than they would if the treatment of labor and capital were uniform. Other common variations in implementation relate to the domestic content benchmark, and whether it is defined as a percentage of physical inputs, sales, costs, or value added.

The effects of domestic content regulations are also determined by firm responses. To begin, it is commonly understood that imports may impose market discipline. If the market that implements domestic content regulations is served by a monopoly input supplier, limits on the use of foreign inputs will reduce competition and allow the domestic input supplier to exploit its market power. In this case, domestic content requirements may even reduce domestic output of the final good, as the domestic input supplier reduces its production and consequent sales to take advantage of its market power. Second, if it is possible for firms to change the scale of their production at low cost, domestic content regulations may fail to shift consumption toward domestic varieties. Finally, if the domestic content requirement is at a level that is too high, foreign firms may reevaluate their decision to supply the domestic market.

The effects of domestic content requirements are also shaped by the nature of the production process. For example, when a content standard is based on value added, whether final good output increases or decreases depends in part on the relationship between labor and the intermediate inputs in production, and whether they are complements or substitutes. Similarly, depending on the relationship between the usage of labor and intermediate inputs, the imposition of a domestic content requirement based on value added may cause the volume of imported intermediate inputs to rise or fall. In addition, the composition of intermediate input imports may change. In particular, if there are many imported intermediate inputs, and the relative price of foreign inputs differs across inputs, a value-added requirement will shift purchases toward foreign inputs whose relative foreign price is the lowest.

In the case of foreign investment, the desire to foster greater domestic activity may be subverted by domestic content regulations. First, since foreign firms are placed at a disadvantage, they may place fewer of their activities in the local market. In addition, when the scale of their operations is reduced, the overall productivity of the foreign operation may suffer from a failure to achieve full economies of scale in production.

Due to the rise of vertically integrated production networks that cross national borders, the proliferation of free trade areas has created new difficulties in defining and meeting content requirements. In particular, when a hub country forms free trade areas with different associates, which are not linked to one another through any common partnership, the definition of content and rules of origin becomes increasingly complicated. The difficulties posed by

multiple partnerships of differing membership groups is acknowledged, for example, by the European Union's creation of the Pan-European Cumulation System in 1997, which integrates bilateral rules into a multilateral framework that enables firms to determine "European" content.

See also rules of origin; trade-related investment measures (TRIMs)

FURTHER READING

Baldwin, Richard. 2006. "Multilateralizing Regionalism: Spaghetti Bowls as Building Blocs on the Path to Global Free Trade." *World Economy* 29 (11): 1451 1518. Addresses political economy forces driving regionalism and how to achieve greater multilateralism, with historical overview.

Corden, W. M. 1971. *The Theory of Protection.* Oxford: Oxford University Press. A key early work on trade policy, including domestic content requirements.

Grossman, Gene M. 1981. "The Theory of Domestic Content Protection and Content Preference." *Quarterly Journal of Economics* 96 (4): 583 603. Determines the effects of content protection, showing it to be equivalent to a combination of other commercial policies.

Krishna, Kala, and Motoshige Itoh. 1988. "Content Protection and Oligopolistic Interactions." *Review of Economic Studies* 55 (1): 107 25. Examines conditions that determine the effects of content protection in oligopolies and thus who might lobby for such protection.

Krueger, Anne O. 1999. "Are Preferential Trading Arrangements Trade Liberalizing or Protectionist?" *Journal of Economic Perspectives* 13 (4): 105 24. Reviews effects of preferential trading areas and finds empirical evidence insufficient.

Lopez de Silanes, Florencio, James R. Markusen, and Thomas F. Rutherford. 1996. "Trade Policy Subtleties with Multinational Firms." *European Economic Review* 40 (8): 1605 27. Shows content requirements can be anticompetitive, reducing output and shifting rents to domestic firms, in oligopolies where foreign multinationals rely more on imported intermediates than do mestic firms (applied to North American auto industry).

Moran, Theodore H. 1998. *Foreign Direct Investment and Development: The New Policy Agenda for Developing Countries and Economies in Transition.* Washington DC: Institute for International Economics. Reviews developing country policies toward foreign firms, including alternatives to domestic content requirements.

Munk, Bernard. 1969. "The Welfare Costs of Content Protection: The Automotive Industry in Latin America." *Journal of Political Economy* 77 (1): 85 98. An early work on domestic content requirements.

DEBORAH L. SWENSON

■ dominant currency

A currency dominates other currencies when it is used more frequently as a unit of account, medium of exchange, and store of value. Local currencies tend to dominate foreign currencies as means of exchange partly because the bulk of transactions is local and partly because governments discriminate in favor of the currency they issue. Legal tender laws and the fact that governments generally disburse funds and receive tax collections in the government-issued currency raise the cost of using other currencies as alternative means of payment in the domestic market. Currency substitution the replacement of domestic money with a (better) foreign money is a slow process even when it is permissible. Inflation must reach high rates before individuals switch out of a currency. For example, an inflation rate of 10 percent a month in the German hyperinflation of the 1920s led to a currency substitution of only 5 percent of the money stock.

At the international level, where legal restrictions matter much less and national currencies can freely compete, specific types of transactions tend to be denominated in international currencies to enhance market transparency, and international investments in different currencies are regularly held as a store of value. In the search for the preferred currency, transaction costs are one of the most relevant considerations. These costs decline as the transaction domain of a currency expands; call this the network value of money. Currencies that are not widely used have a low network value and are at a competitive disadvantage against widely used currencies. Once a currency reaches a dominant network value, it may

gain from inertia: newer currencies with similarly low transaction costs may not be able to quickly upstage the dominant currency (Rey 2001).

Historical evidence indicates that one currency tends to dominate others as an international medium of exchange and a store of value. The Roman silver *denarius* was the first world currency; the Byzantine *solidus* was the unchallenged coin from the 5th to the 7th centuries. But the international role of the *solidus* was challenged by the Islamic *dinar*, which eventually made the crossover; both lasted until the 12th century. In the 13th century, Italian coins came to prominence: the Genoese *genoino*, the Florentine *fiorino*, and the Venetian *ducato*. All three coins circulated side by side for quite some time (Cipolla 1956). In the 19th century, Britain was the leading industrial economy in the world, and the British pound became the leading international currency. Britain's economic preeminence came to an end after World War I, but the key status of the pound lasted for more than four more decades (Eichengreen 2005). As late as 1965, 20 percent of official reserves were denominated in pounds, demonstrating how slowly a dominant currency can decline. The U.S. dollar emerged as the dominant international currency after World War II, but lost some ground toward the end of the 20th century, first with respect to the deutsche mark and the Japanese yen and later to the euro.

The U.S. Dollar and the Euro as Dominant Currencies Several different functions are subsumed under the label of "international role of a currency": to invoice internationally traded goods and services, to employ as a vehicle currency (that is, as neither the exporter's nor the importer's currency) in foreign exchange markets, to denominate assets held by monetary authorities and the private sector, and to serve as a reference currency in defining fixed exchange rates (Cohen 1971; Kenen 1983).

The dollar remains the dominant invoice currency, even though the extent of its dominance has declined since the advent of the euro (Goldberg and Tille 2006). Specifically, according to available data,

the U.S. dollar remains the dominant invoice currency outside of Europe, where the euro has surpassed the dollar. The dollar is the largest currency traded in the foreign exchange markets, with a market share of 45 percent of daily transactions; the euro follows with a share of 20 percent (BIS 2005). These shares remained stable between 2001 and 2004. The greatest advances made by the euro as an international currency are in international bond issues. According to data published by the Bank for International Settlements (2006, table 13 B), international bonds and notes denominated in euros exceeded those denominated in dollars. The ascendancy of the euro coincides with the increased degree of efficiency, liquidity, and integration of the euro financial markets (Portes and Rey 1998).

Following the depreciation of the dollar relative to the euro after 2002, the financial press focused on the prospect that central banks with large and increasing stock of international reserves   especially those in Asia   may want to substantially diversify their holdings out of dollars and into euros and, in the process, bring an end to the dominance of the dollar in official portfolios. Yet, according to data published by the International Monetary Fund (2005, table I.2), the dollar still retained the same reserve share in 2005 that prevailed at the end of the Bretton Woods system. The novel aspect in the data is that the euro has gained at the expense of currencies other than the dollar (the euro share in official reserves has gone from 6.7 percent of the combined shares of the legacy currencies mark, franc, and guilder in 1973 to 25 percent in 2004). It is safe to say that the euro is becoming an established alternative to the dollar (Chinn and Frankel 2005).

With the United States and the euro area converging to similar economic and financial size, the race is now between two currencies that are backed by large and diversified regions and are integrated with the rest of the world both in trade and finance. Differences in policies will determine the outcome of the match for dominance. Hence the concern that the decline of the dollar may pick up speed if fiscal policy in Washington does not change.

Benefits of Currency Dominance The United States benefits from the dollar as the preeminent international currency in a number of ways. First, it earns a "foreign" seigniorage on the dollar currency held abroad; data indicate that as of 2005, $352 billion, approximately 50 percent of outstanding U.S. dollar currency, is in the hands of nonresidents (Survey of Current Business 2006). Second, the special role of the dollar has permitted the United States to issue liquid and short-term foreign liabilities and invest in illiquid and long-term foreign assets; that is, the United States is the "banker of the world" (Despres et al. 1966). By borrowing cheaply relative to rates of return achieved on foreign assets, the United States has relaxed the external constraint typically faced by other countries. This is one reason that, as of 2007, the United States had accumulated a staggering net foreign debt without yet incurring a financial crisis.

Finally, U.S. importers are largely insulated from exchange rate movements because virtually all U.S. imports are invoiced in dollars, and only a small fraction of a change in the exchange rate is passed through import prices. This reduction in exchange rate risk means that a depreciation of the dollar does not have the traditional effect of significantly reducing the demand for imports and does not lead to a deterioration of the terms of trade. The policy consequence of this benefit to U.S. consumers is that a dollar depreciation is a weak tool to produce a switch from imports to domestic goods and services and that a reduction of the trade deficit occurs mostly through the export side.

In sum, the dollar remains the dominant international currency; the euro is a distant second most important international currency. Whether the euro will eventually overtake the dollar depends on the inertia accorded to the dominant currency and on the policies that will be pursued on both sides of the Atlantic.

See also Bretton Woods system; common currency; convertibility; currency competition; currency substitution and dollarization; dollar standard; euro; exchange rate pass-through; exorbitant privilege; gold standard, international; multiple currencies; optimum currency area (OCA) theory; reserve currency; seigniorage; vehicle currency

FURTHER READING

Bank for International Settlements. 2005. *Triennial Central Bank Survey: Foreign Exchange and Derivative Market Activity in 2004*. Basel, Switzerland: Bank for International Settlements. Provides data on the size and currency breakdown of the foreign exchange and derivative markets.

———. 2006. *BIS Quarterly Review* (June). Gives developments in international banking and international financial markets.

Chinn, Menzie, and Jeffrey Frankel. 2005. "Will the Euro Eventually Surpass the Dollar as Leading International Reserve Currency?" NBER Working Paper No. 11510. Cambridge, MA: National Bureau of Economic Research. Analyzes the rise of the euro as an international currency and as a competitor of the U.S. dollar.

Cipolla, Carlo M. 1956. *Money, Prices, and Civilization in the Mediterranean World*. Princeton, NJ: Princeton University Press. Gives a historical account of early and dominant moneys.

Cohen, Benjamin J. 1971. *The Future of Sterling as an International Currency*. London: Macmillan. The first article to define the functions of an international currency.

Despres, Emile, Charles Kindleberger, and Walter Salant. 1966. "The Dollar and World Liquidity: A Minority View." *The Economist* (February 5). Advances the thesis that the United States acts as the world's banker, that is, borrows short and lends long.

Eichengreen, Barry. 2005. "Sterling's Past, Dollar's Future: Historical Perspectives on Reserve Currency Competition." NBER Working Paper No. 11336. Cambridge, MA: National Bureau of Economic Research. Underscores that it took a relatively long time for the UK pound to lose its status of key currency.

Goldberg, Linda, and Cédric Tille. 2006. "The International Role of the Dollar and Trade Balance Adjustment." NBER Working Paper No. 12495. Cambridge, MA: National Bureau of Economic Research. Quantitative study on the dominant role of the U.S. dollar in international transactions.

International Monetary Fund. 2005. *Annual Report of the Executive Board for the Financial Year Ended April 30, 2005*. Washington, DC.

Kenen, Peter B. 1983. *The Role of the Dollar as an International Currency*. New York: Group of Thirty. Elaborates on the different functions of an international currency.

Nguyen, Elena A. 2006. "The International Investment Position of the United States at Yearend 2005." *Survey of Current Business* 86 (7) (July): 9 19.

Portes, Richard, and Hèléne Rey. 1998. "The Emergence of the Euro as an International Currency." *Economic Policy* 26: 305 43. An early article advancing the thesis that the euro, the new European currency, can become a competitor of the U.S. dollar in international transactions.

Rey, Hèléne. 2001. "International Trade and Currency Exchange." *Review of Economic Studies* 68 (2): 443 64. Links the emergence of vehicle currencies to the pattern of international trade; also argues that inertia plays a role in the rise and fall of dominant monies.

MICHELE FRATIANNI

■ dual exchange rate

The term *dual exchange rate* refers to a situation in which more than one exchange rate applies between one currency and another. The term arises most often when a country's authorities establish one exchange rate for certain transactions involving foreign exchange/currency and a second rate governing other transactions. Many countries in Europe and the developing world use a fixed exchange rate for commercial (current account) transactions and another rate, either fixed or floating, for other (e.g., financial account) transactions. Another type of dual exchange rate occurs when a country's authorities require certain foreign exchange transactions to take place at a prescribed exchange rate, leaving other transactions to take place at the rate prevailing in the commercial market or driving other transactions into an illegal ("black") market, again at a more expensive (i.e., more depreciated) exchange rate.

For example, some countries with currencies that are not freely convertible have required exporters to sell ("surrender") their foreign exchange earnings to the country's central bank at one exchange rate and then buy the foreign exchange they need in the commercial market at the prevailing, and probably more depreciated, market clearing rate. In some other countries, tourists must exchange a certain amount of foreign currency at a set exchange rate on entering the country, the rate being different from, and usually more appreciated than, that in the commercial market. When the second market is illegal, the difference between the official rate and that in the illegal market is called the black market premium.

Dual exchange rates have been created for various reasons. Many European and Latin American countries, for example, have used dual rate regimes to ease the transition from a fixed to a floating rate regime or to limit downward pressure from capital flight on the rate for current transactions (Marion 1994). In some centrally planned or developing economies with nonconvertible exchange rates, they have been created to support activities or sectors said to yield special benefits to the economy. Authorities have sometimes argued that certain imports   goods used in the manufacture of exports or imports for official purposes such as development projects   are more valuable than imports of consumer goods and thus deserve priority, including access to cheaper foreign exchange. Accordingly, they have allowed importers of such goods easier access to foreign exchange at the official exchange rate, or allowed them to purchase foreign currency at a less expensive exchange rate. Dual exchange rates of this type have often been linked to systems of import licenses, in which importers request permission from the authorities to obtain foreign exchange, specifying the purpose(s) for which the foreign exchange will be used.

Despite their political appeal, dual exchange rate systems introduce serious inefficiencies to the foreign exchange market. Dual exchange rate systems artificially segment the market for foreign exchange. When some activities obtain foreign exchange at less expensive rates while others must pay more, the result is a nontransparent system of subsidies to favored

sectors and taxes on those disfavored. This system distorts the allocation of resources and lowers output and real economic growth, as foreign exchange no longer flows to the most productive activities.

Dual exchange rate systems also create incentives to engage in wasteful activities (i.e., rent-seeking activities) as a way of obtaining foreign exchange at preferential rates. Importers, for example, may shift the composition of their imports or engage in illicit transactions to obtain a more favorable foreign exchange rate (Gros 1987). Exporters may engage in underinvoicing as a way to keep foreign exchange out of the country or to avoid having to "surrender" their foreign exchange earnings. Finally, dual exchange rate systems enable those in favored sectors to operate more cheaply, gain advantages relative to other sectors, or sell foreign exchange to those in less favored sectors, thus obtaining profits at the expense of others.

Because of the distortions created by dual exchange rate systems, economists typically discourage their adoption, and dual exchange rates have become far less common since the end of the 1980s. If national authorities wish to encourage certain activities, a more transparent system of subsidies and taxes has a less disruptive effect on the pricing signals created by exchange markets.

See also balance of payments; black market premium; capital controls; capital flight; convertibility; debt defla-tion; exchange rate regimes; foreign exchange intervention; peso problem

FURTHER READING

Gros, Daniel. 1987. "Dual Exchange Rates in the Presence of Incomplete Market Separation: Long Run Effec tiveness and Implications for Monetary Policy." IMF Working Paper No. 87/45 (July). Washington, DC: International Monetary Fund. Discusses the impact of dual exchange rates on domestic monetary policy.

Lai, Ching Chong, and Yun Peng Chu. 1986. "Exchange Rates Dynamics under Dual Floating Exchange Rate Regimes." *Southern Economic Journal* 53 (2): 502 8. Argues that eliminating a dual floating exchange rate system may trigger an overadjustment ("overshooting") of the primary (e.g., commercial) market exchange rate, so countries are urged to unify dual rates under appro priate safeguards, such as a (temporary) tightening of monetary policy.

Marion, Nancy P. 1994. "Dual Exchange Rates in Europe and Latin America." *World Bank Economic Review* 8: 213 45. Discusses the reasons why dual exchange rates have been created in European and Latin American countries.

JOSHUA GREENE

■ dumping

See anti-dumping

■ early warning systems

Early warning systems (EWSs) are used by policymakers, ratings agencies, and many financial market participants to assess the vulnerability of countries to currency crises. Emerging markets have been in the eye of the storm several times in the late 20th and early 21st centuries: Mexico in 1994, Southeast Asia in 1997, Russia in 1998, and Argentina in 2001, among others. The effects of some of these episodes were not confined to the countries themselves but spread to other economies as well. In addition, the speculative attacks against currencies participating in the European Monetary System in the 1980s and the early 1990s showed that developed economies are not immune from such events either.

The scope, spread, and severity of crises originating in the foreign exchange market highlight the importance of understanding their causes and predicting where and when they may happen. This need cannot be overstated in an increasingly integrated world economy where free-flowing capital exposes the world's markets to greater risks. The so-called first-, second-, and third-generation models of currency crises constitute the theoretical framework for explaining the mechanisms underlying episodes of increased speculative pressure in the foreign exchange market. EWSs empirically test the implications of these models with the aim to predict currency crises and inform the relevant policy actions.

Basic Components of an EWS The first challenge in the construction of an EWS is how to define a crisis from an empirical perspective. A variety of methods have been used to identify such episodes, ranging from straightforward measures of currency depreciation (e.g., Frankel and Rose 1996) to composite indexes incorporating changes in exchange rates, interest rates, and international reserves (e.g., Eichengreen et al. 1995). Measures focusing exclusively on developments in exchange rates can capture only the effect of successful speculative attacks, whereas wider market pressure indexes may, in addition, encapsulate failed attempts to destabilize an exchange rate regime or put downward pressure on a currency. This follows from the fact that policymakers may choose to use the interest rate or part of the stock of the country's international reserves in order to prevent a devaluation of the currency. If such intervention is successful, then the movement in the exchange rate may be limited (or even zero), in which case a simple measure based solely on exchange rate movements will not register the foreign exchange market activity. In contrast, a wider pressure index will capture such activity through the changes in the international reserves and/or the interest rate components of the index.

A second issue relating to the construction of an EWS concerns the choice of sample, data frequency, and variables to be included in the model. Samples can be limited to economies that share certain characteristics (e.g., the stage of their development or their geographical proximity, among other things) or can be extended to include all economies for which data are available. There is a trade-off between the need to avoid overaggregation and the need to maximize the number of observations available for analysis.

An interrelated issue is that of the frequency of the data. Higher-frequency data (e.g., monthly) will provide more degrees of freedom but several economic variables are observed at lower frequencies (e.g., annual). Hence, adopting a monthly frequency increases the number of observations for the available variables but precludes the use of variables that are observed at lower frequencies. Adopting an annual frequency allows the use of a greater number of variables but reduces the number of observations per variable. Furthermore, selecting an annual frequency may have implications for identifying crisis episodes, as the effects of a successful or unsuccessful speculative attack may have played out within the twelve-month period. In such cases, annual data will miss the crises altogether.

Also, a choice must be made of the variables or indicators to be included in the EWS. The selection is usually informed by the relevant theory. First-generation models stress the importance of fiscal and monetary indicators (e.g., the fiscal deficit, the change in the amount of money in the economy); second-generation models emphasize the role of variables related to market expectations (e.g., the interest rate differential with respect to the United States); and third-generation models focus on, among others, the need to monitor indicators of the health of the banking system (e.g., the ratio of non-performing loans to performing ones).

The decisions on how to measure a crisis, and what sample, frequency, and variables/indicators to use constitute one reason for the different types of EWSs that have surfaced in the literature. The other reason concerns the distinct methodologies employed, which range from signal extraction to estimating discrete binary choice and structural models.

The "Signals" Approach One of the most influential approaches in EWSs is the "signals" approach, pioneered by the economists Graciela Kaminsky, Saul Lizondo, and Carmen M. Reinhart (see Kaminsky et al. 1998). This involves monitoring a set of key economic variables over time. Whenever a value of a variable exceeds (or, depending on the nature of the variable, falls below) a predefined critical threshold the model issues a warning signal that a

crisis may be forthcoming within a specified time period. In order to assess the predictive ability of the model the incidence of signals needs to be compared with the incidence of actual crisis episodes.

Kaminsky et al. (1998) construct an exchange market pressure index using a weighted scheme involving changes in the exchange rate and the stock of international reserves. This is in turn converted into a binary crisis variable by assigning a value of one whenever a value exceeds the mean of the index by more than three standard deviations and assigning a value of zero otherwise. In that way, the ones correspond to crisis incidents and the zeros correspond to normal periods.

The critical thresholds for a signal to be issued are chosen on the basis of minimizing the ratio of "bad" to "good" signals, where a bad signal is one that is not followed by an actual crisis and a good signal is one that is followed by an actual crisis. Kaminsky et al. set a signaling horizon of 24 months, that is, if a signal has been issued that is followed by a crisis episode within this time window it will be considered a good signal. This is one of four possible outcomes. A second possible outcome is that a signal has been issued with no crisis occurring within the 24-month period. A third one is that a signal has not been issued but a crisis does take place within the specified time window. The final possible outcome is that a signal has not been issued and no crisis takes place.

We can summarize this information in a matrix. A is the number of months that a good signal was issued, B is the number of months that a bad signal was issued, C is the number of months that a signal should have been issued, and D is the number of months that a signal was, rightly, not issued. In terms

	Crisis (within 24 months)	No crisis (within 24 months)
Signal was issued	A	B
No signal was issued	C	D

Source: Kaminsky et al. (1998).

of this matrix, optimal thresholds are chosen so that [B/(B+D)]/[A/(A+C)], the "noise-to-signal" ratio, is minimized. These thresholds are expressed relative to the percentiles of an indicator's distribution. Several threshold levels are examined in the vicinity between 10 percent and 20 percent and the one minimizing the noise-to-signal ratio is chosen as the optimal threshold. This is then used as a reference level for all countries in the sample (of course, a given percentile will correspond to different actual threshold values in different countries).

Kaminsky et al. use 15 indicators classified in the following categories: capital account, current account, debt profile, international variables, financial liberalization, other financial variables, real sector, fiscal variables, institutional/structural factors, and political variables. They find that a range of indicators contain useful information, including the deviation of real exchange rate from trend, banking crises, exports, equity prices, the ratio of M2 to international reserves, output, excess M1 balances, international reserves, the M2 multiplier, the ratio of domestic credit to gross domestic product, the real interest rate, and the terms of trade. These results have been partly replicated using a different sample of countries and time period by others (e.g., Berg and Patillo 1999).

The Discrete-Choice Approach Another type of EWS uses probit (or logit) econometric estimations to calculate the probability of a crisis conditional on a set of variables. The dependent variable is a binary indicator with unity indicating a crisis observation and zero indicating a noncrisis observation. The main advantage of this approach is that a variety of potential "triggers" can be used as explanatory variables of crisis episodes. For example, the predictive ability of economic and structural variables corresponding to the first-, second-, and third-generation theoretical models of currency crises can be assessed in this framework.

An example of the discrete-choice approach is Frankel and Rose (1996). They use annual observations for 105 countries between 1971 and 1992. Controlling for a range of macroeconomic indicators, they focus on the composition of debt and capital flows. They define a crisis as a reduction in the value of the domestic currency of at least 25 percent that also exceeds the previous year's depreciation by at least 10 percent. This measure would not capture a failed speculative attack as the crisis definition does not include measures of international reserves and/or the interest rate. Using a probit model, the authors find that crisis probabilities are significantly influenced by developments in foreign direct investment, international reserve levels, domestic credit, U.S. interest rates, and the real exchange rate.

Large sample studies can be useful in identifying potential crisis-triggering factors. Pooling observations from regions with different economic, historical, and structural characteristics, however, calls into question the general applicability of the model's implications: it may be necessary to focus on individual crisis incidents so that valuable information, which may be specific to an incident, does not go amiss. The following type of model is dedicated to a more structural approach.

The Structural Approach This type of model is used to examine a country's vulnerability to a currency crisis by assessing certain economic characteristics. An example of the structural approach is Sachs et al. (1996). These authors construct a market pressure index involving the exchange rate and international reserves but, unlike the signals and the discrete-choice approaches, they do not convert it into a binary variable. Instead, they use the index as the dependent variable in a regression. The right-hand side of the equation includes a measure of real exchange rate misalignment, a measure of liquidity in the economy, and a set of dummies capturing weaknesses in fundamentals and low international reserves. The sample consists of a cross-section of 20 emerging economies, and the time span for the calculation of the variables concentrates on the time around the Mexican crisis of 1994. Even though the results reveal that the indicators in question have explanatory ability with respect to the effects of pressure in the foreign exchange market on financial crises, subsequent efforts by Berg and Patillo (1999) to forecast the 1997 Asian crisis using this approach were not successful.

The Way Forward The definition of a crisis, the methodology employed, sample selection, and other factors can affect the predictive ability of the three main types of EWSs. But despite their limitations, EWSs of financial crises are a useful tool for policymakers. Although there is no consensus in the literature about a definitive list of predictors of crisis episodes, certain indicators, such as the misalignment of the real exchange rate and excess liquidity, are important factors in several models.

This is a literature that is constantly growing. Recent advances in econometrics have resulted in the construction of increasingly sophisticated models. The challenge for researchers and policymakers is that different crisis episodes appear to have different causes. It remains to be seen whether our current knowledge will be sufficient to enable us to predict the next crisis.

See also banking crisis; contagion; currency crisis; equilibrium exchange rate; European Monetary Union; exchange market pressure; financial crisis; financial liberalization; foreign exchange intervention; international reserves; money supply; real exchange rate; spillovers; sterilization

FURTHER READING

Berg, Andrew, and Catherine Patillo. 1999. "Are Currency Crises Predictable? A Test." *IMF Staff Papers* 46 (2): 107 38. Evaluates the out of sample predictive ability of models constructed prior to the 1997 Asian crisis.

Eichengreen, Barry, Andrew Rose, and Charles Wyplosz. 1995. "Exchange Market Mayhem: The Antecedents and Aftermath of Speculative Attacks." *Economic Policy* 21: 251 98. Employed for the first time a market pres sure index, which in addition to the usual exchange rates and international reserves variables included interest rate differentials.

Frankel, Jeffrey, and Andrew Rose. 1996. "Currency Crashes in Emerging Markets: An Empirical Treatment." *Journal of International Economics* 41: 351 66. An ex ample of a large sample EWS based on probit analysis.

Kaminsky, Graciela, Saul Lizondo, and Carmen M. Rein hart. 1998. "Leading Indicators of Currency Crises." *IMF Staff Papers* 45 (1): 1 48. Regarded as the seminal work on signal based EWSs and generated a number of citations and replications.

Sachs, Jeffrey, Aaron Tornell, and Andrés Velasco. 1996. "Financial Crises in Emerging Markets: The Lessons from 1995." *Brookings Papers on Economic Activity* 1: 147 215. Focuses on the 1994 Mexican crisis and adopts a structural approach to examine the vulnerability of countries to speculative pressures.

ALEX MANDILARAS

■ Economic Community of West African States (ECOWAS)

The Economic Community of West African States (ECOWAS), also known simply as the community, is a regional trade organization comprising countries in the West African region. As of late 2007, Benin, Burkina Faso, Cape Verde, Côte d'Ivoire, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mali, Niger, Nigeria, Senegal, Sierra Leone, and Togo (after Mauritania withdrew in 1999) are members. The aims and objectives of the community are to promote trade, cooperation, and self-reliance in West Africa and to assist in the creation of the African Economic Community (AEC).

Origin and Background Like other regional economic groups in Africa, the community was formed based on the pan-African vision of economic integration, which gained prominence during the late 1950s and early 1960s. It was believed that the small and fragmented postcolonial national markets would constitute an obstacle to the economic development of Africa. The pan-African solution was to create a single common market for the continent. The pan-African approach to integration was a "bottom-up" process that would begin at the regional levels (central, east, north, south, and the west) and evolve into a continentwide common market by welding the regional organizations together. It was against this background that the community was created to overcome the isolation of the small West African countries following the colonial and postindependence nationalism periods.

The community was created by treaty in 1975 as a preferential trade area for the transformation of the West African region. The original plan for the community was to establish a free trade area by 1980 followed by a customs union by 2000, and an economic and monetary union by 2005. Owing to a number of factors, however, including internal sociopolitical and economic instability in member countries, poor coordination of macroeconomic policies at the regional level, weaknesses in the operational procedures of the community's institutions, and lack of commitment on the part of member countries to implement the community's priority programs, implementation of the integration programs has proceeded very slowly.

The ECOWAS Treaty was revised in 1993 to expedite the process of economic integration and to increase political cooperation in the region. The revised treaty articulates comprehensive cooperation and integration programs, including trade promotion and liberalization; the provision of better roads and telecommunications infrastructure; and the development of agriculture, industry, and the energy sector. The implementation of the trade liberalization programs was planned to occur in three stages. The first involves the full liberalization of trade in unprocessed goods and traditional handicrafts. The second entails a phased liberalization of trade in industrial products whereby the most advanced member countries (Nigeria, Ghana, and Côte d'Ivoire) were to remove all trade barriers in a period of six years, middle-group countries (Benin, Guinea, Liberia, Sierra Leone, and Togo) within eight years, and third-group countries within ten years. The third stage concerns the establishment of a common external tariff.

In order to accelerate the regional integration process, the revised treaty allows for a double-track (fast-track) approach to economic integration whereby a group of countries can take measures to accelerate integration among themselves. The revised treaty also assigned to the community the additional responsibility of preventing and settling regional conflicts.

Various institutions have been formed to facilitate the achievement of the broad aims and objectives of the community. These institutions include the decision-making organs responsible for the smooth functioning of the community, the Mediation and Security Council (for maintaining peace and security in the region), and the Fund for Cooperation, Compensation, and Development (which provides development capital). The decision-making organs include the Authority of Heads of State and Government (i.e., the supreme policy organ), responsible for the general direction and control of the community; the Council of Ministers, responsible for making decisions on programs and the functioning and development of the community; the Parliament (without legislative powers), which acts in a consultative and advisory capacity; the Economic and Social Council, which plays an advisory role; the Court of Justice, which interprets the provisions of the ECOWAS Treaty and settles disputes among member states that are referred to it; and the Executive Secretariat, responsible for the smooth functioning of the community and for the implementation of the decisions of the supreme policy organ. The community also has specialized technical commissions that prepare projects and programs for consideration by the Council of Ministers on all key sectors of the economy.

The member states of the community have diverse backgrounds. Their heritages vary from Anglophone to Francophone states and to independent Liberia. The region consists of mainly two geographical zones the Sahelian, semiarid, largely landlocked zone, and a more humid, forested coastal zone. Nigeria is a dominant economy in the region, accounting for more than half of the population and gross domestic product (GDP) of the region. Most of the other states are small in size (less than 5 percent) in terms of population and GDP. The region's resource endowments consist mainly of crude oil (in Nigeria) and agricultural products such as cocoa beans, coffee, and yams. The per capita GDP varies widely across the region. For example, in 2004, it was U.S. $141 for Liberia and U.S. $1,979 for Cape

Verde. In the same year, the UN Development Program ranked only Cape Verde and Ghana as medium human development countries and the rest as low human development countries.

Key Elements and Procedures Since the early 1990s, there has been some progress toward the establishment of a free trade area and a common market. For instance, visa and entry-permit requirements for ECOWAS nationals have been abolished and an ECOWAS passport has been adopted. In addition, ten countries (Benin, Gambia, Ghana, Guinea, Mali, Niger, Nigeria, Senegal, Sierra Leone, and Togo) have removed barriers on trade in unprocessed goods and three (Benin, Gambia, and Sierra Leone) have eliminated barriers on trade in industrial products. All member countries have abolished monetary nontariff barriers. Nonmonetary nontariff barriers still exist, however. A common external tariff was scheduled to be in place by the end of 2007.

In 1994, eight of the countries (Benin, Burkina Faso, Côte d'Ivoire, Guinea Bissau, Mali, Niger, Senegal, and Togo) took advantage of the double-track innovation and came together to create a customs union and a monetary zone called the West African Economic and Monetary Union (WAEMU). The WAEMU member states implemented a common external tariff in 1998 and have started to harmonize their policies in various fields. WAEMU has generally been perceived as successful. Intra-WAEMU trade as a share of the region's total world trade increased from about 10 percent in 1994 to about 16 percent in 2000. Also, the average annual GDP growth rate for the WAEMU subregion has been higher than that of sub-Saharan Africa as a whole. Six other countries of the community, Gambia, Ghana, Guinea, Liberia, Nigeria, and Sierra Leone, have followed suit and embarked on an initiative to form a second monetary zone with the aim to merge with the WAEMU and eventually create a single monetary zone for the ECOWAS region.

Until now, the community has operated as a preferential trade area. The member states apply tariff preferences under the common preferential tariff on goods originating from the community and charge national tariff rates on goods originating from nonmember countries. Member countries use the community's rules of origin to determine the tariff rate to levy on imported products. Products are deemed to originate from the community if (1) they are wholly obtained within a member state, (2) they have undergone sufficient processing in a member state (usually 40 percent value added), (3) 60 percent of the raw material used to manufacture the product comes from member states, or (4) they are local products or traditional handicrafts. Normally, a certificate of origin is not required for local products and traditional handicrafts of ECOWAS origin. For any other product, however, an importer who seeks to benefit from the preferential treatment must provide a certificate of origin.

Impact on Member States Regional economic integration agreements can be beneficial to member countries in both the static (short-run) and the dynamic (long-run) sense. In his analysis of static effects, Jacob Viner (1950) observed that while regional economic integration represents a movement toward free trade on the part of member countries (leading to trade creation), it can also lead to diversion of trade from low-cost nonmember sources that face tariffs to a high-cost member source that no longer faces tariffs. In the short run, member states can benefit as long as the trade creation effects exceed the trade diversion effects. Using the gravity framework, Jacob W. Musila (2005) estimated the two static effects for ECOWAS and found that trade creation exceeds the trade diversion effects. Therefore, member countries of the ECOWAS organization can be expected to obtain net gains in welfare. Céline Carrére (2004) also found that the formation of ECOWAS is associated with the increase in intra-ECOWAS trade and that the currency union, WAEMU, has helped to reduce trade diversion effects. The performance of intra-ECOWAS trade shows that its share in the region's total world trade increased from 2.4 percent in 1975 to more than 10 percent in the 1990s. During the period 1994 98, owing to the improvement of performance in the WAEMU zone, the share of intra-ECOWAS trade increased slightly but steadily from 10.7 to 11.9 percent.

The dynamic benefits of a regional integration occur through well-known channels, namely, increased efficiency through specialization, economies of scale, or increased trade and investment. Like static effects, however, the net impact of dynamic effects is not obvious. In the case of ECOWAS, exports comprise a limited range of agricultural commodities (only Nigeria is a net oil exporter), and manufactured goods are imported mainly from nonmember countries. Such pattern of trade can be a hindrance to long-term economic growth of the region since agricultural products are more prone to deterioration in terms of trade. Indeed, the sluggish growth that characterized the ECOWAS region in the late 1990s was partly attributed to the sharp fall in cocoa prices. Augustin Fosu (1990) finds evidence that supports the view that exporting primary commodities does not significantly improve a country's long-term economic growth rate. Antonio Spilimbergo (2000) shows that the importation of manufactured goods from rich countries (the North) by poor countries (the South) may not guarantee economic growth in the South. The learning-by-importing models, however, suggest that imports of high-technology goods can result in the transfer of technology which, in turn, stimulates domestic innovation and economic growth in the importing country.

Relation with External Actors The relation between ECOWAS and external actors in the world economy varies from region to region. ECOWAS and other African regional blocs seek to augment and deepen economic integration in their respective regions with a view to forming an AEC that will enable Africa to be in a stronger position when negotiating trade matters with other regional organizations such as the European Union (EU), the South American regional economic organization Mercosur, and the North American Free Trade Area. ECOWAS does not have overlapping memberships, which have been seen as an obstacle to market enlargement in some African regional integrations.

On the other hand, the relation between ECOWAS and non-African trade blocs is viewed by some analysts to be an impediment to regional integration. Jeffrey D. Lewis, Sherman Robinson, and Karen Thierfelder (2003) have argued that North-South trade is more attractive to most African countries than South-South trade because of the structure of their economies or colonial heritage. Indeed, ECOWAS member states trade with the EU and the United States more than with one another. This trade is set to increase even more following the opening up of the EU and U.S. markets under the Everything-but-Arms and the African Growth and Opportunity Act initiatives, respectively. This may work to further reduce the relative importance of intra-ECOWAS trade. In addition, the World Trade Organization provision that substantially reduces quotas and duties on goods originating from the world's poorest countries will also increase the region's trade with non-ECOWAS countries and further reduce the advantages that the community offers to its member states. The very processes of global trade liberalization and cooperation with non-African trade blocs have contributed to the evolution of ECOWAS, however. To reduce transaction costs, the EU, for example, prefers developing countries to negotiate as one group (instead of as numerous individual countries). Furthermore, the success of the EU also inspires the evolution of ECOWAS. Indeed, the ECOWAS institutional design is fashioned after the EU.

See also customs unions; free trade area; regionalism; rules of origin

FURTHER READING

ECOWAS Treaty. 1975. Downloadable from http://www.ecowas.int/. Gives detailed insights about the establishment of ECOWAS, its aims and objectives, and the various articles and protocols about the operation of the organization.

Carrére, Céline. 2004. "African Regional Agreements: Impact on Trade with or without Currency Union." *Journal of African Economies* 13 (2): 199 239. Examines the impact of regional agreements on intraregional trade and on trade with the rest of the world, and compares the effects of the preferential trade agreements and currency unions in Africa.

Fosu, Augustin K. 1990. "Export Composition and the Impact of Exports on Economic Growth of Developing

Economies." *Economics Letters* 34 (1): 67 71. An em pirical investigation suggesting that exporting primary products does not have a significant influence on long run economic growth rates of the producing country.

Lewis, Jeffrey D., Sherman Robinson, and Karen Thier felder. 2003. "Free Trade Agreements and the SADC Economies." *Journal of African Economies* 12 (2): 156 206. A general equilibrium analysis of the impact of trade liberalization in the Southern African Develop ment Community (SADC) that shows that trade with the EU is more beneficial for SADC countries than a SADC free trade area.

Musila, Jacob W. 2005. "The Intensity of Trade Creation and Trade Diversion in COMESA, ECCAS and ECOWAS: A Comparative Analysis." *Journal of African Economies* 14 (1): 117 41. An estimation of the static gains/losses of selected regional economic integration in Africa using the gravity model.

Spilimbergo, Antonio. 2000. "Growth and Trade: The North Can Lose." *Journal of Economic Growth* 5 (2): 131 46. An analysis of North South trade that suggests that rich nations (North) may experience technological slowdown if they trade with poor nations (South); and the South could also lose in such trade.

Viner, Jacob. 1950. *The Customs Union Issue*. New York: Carnegie Endowment for International Peace. Examines the effects of regional trade agreements and introduces the concepts of trade creation and trade diversion for the first time.

JACOB W. MUSILA

■ economic development

Economic development consists of socioeconomic changes that have the goal of raising economic wel fare and the standard of living of a population. Chief among these objectives are higher income per capita and lower poverty, but others involve increasing protection of the natural environment, health and life expectancy, literacy and the educational attain ment of current and future generations, and political freedom and civil liberties.

There are four major, proximate forces that in fluence economic development: population growth and demographic change; investment and physical capital accumulation; human capital and techno logical change; and land, agriculture, and geography. Underlying how these forces affect development, however, are policy choices. Three main areas of pol icy debates are discussed below: trade liberalization, international financial liberalization, and foreign aid. But these policies have underlying determinants as well. Public sector institutions and governance are influenced by political economy, social conflict, and the presence of large rents, among others. All of these forces have international linkages. The next sections highlight these connections of development to the world economy.

Population, Demographic Change, and International Migration The population in the devel oping world rose from 2.325 billion in 1960 to 5.489 billion in 2006. From Thomas Malthus to Robert Solow, a popular body of work has grown postulating how higher rates of population growth are linked to lower standards of living. Malthus focused on the impact of population growth on scarce food supplies. Solow postulated that increased population growth causes the amount of capital per worker to decline, reducing productivity and income per capita. Initial testing of Solow's hypothesis, such as that produced by Mankiw, Romer, and Weil (1992), showed that, holding other things constant, higher rates of pop ulation growth were indeed connected to lower in come per capita.

But a more nuanced view of the relationship has emerged from recent research, one that is mediated by the economy's dependency rate, defined as the sum of people less than 15 years old and those older than 64, divided by those aged 15 to 64. As popu lation growth expands, often as a result of the so called demographic transition, dependency rates first rise in response to the larger proportion of children in the population. But as this wave of chil dren becomes adults and enters the labor force it can lead to a surge in the saving rate, capital accu mulation, and productivity growth. The result is a period of increased economic growth. Bloom and Williamson (1998) find this demographic dividend to be important in explaining the rapid growth of the

East Asian economies in the period 1960 to 1990, accounting for 1.9 percentage points out of the 6.9 percent average rate of growth of these economies. As the population ages, however, dependency rates increase and growth is negatively affected.

Population flows across countries have exploded in recent years and international migration has become a significant development force. In 1960, there was a stock of slightly over 75 million people residing in countries other than their country of birth. By 2005, this number had expanded to 190 million. Most of these migrants 65 percent of them were born in developing countries.

The impact of emigration on developing countries has been a matter of debate. A partial-equilibrium approach would suggest that emigration reduces labor supply and raises wages in source countries. But in the Heckscher-Ohlin-Samuelson (HOS) general equilibrium framework used in trade theory, emigration has no lasting impact. As migrants leave and wages rise, production shifts away from labor-intensive products and into capital-intensive sectors. The overall demand for labor in the economy shrinks, matching the reduced labor supply induced by emigration (Rivera-Batiz 1983). But the assumptions of the HOS model under which this result is derived are stringent. Once these assumptions are relaxed, the theoretical analysis may yield complex and ambiguous effects of emigration on source countries (Rivera-Batiz 2008). The empirical evidence on the issue has grown in recent years, but remains exploratory (see, for example, Mishra 2006). In some countries, such as China, periods of emigration have been followed by a rise of reverse migration, which returns workers with accumulated skills to the source countries.

The impact of emigration in developing countries is often dominated by the flow of remittances. Measured in international purchasing power parity dollars, developing countries received more than $600 billion in 2005, and for some countries, such as Bangladesh, India, Mexico, Pakistan, and the Philippines, remittances are a major item in the balance of payments. Remittances raise households' standard of living, reduce poverty, and contribute to devel-

opment. But whether they simply raise current consumption instead of stimulating investment and future growth varies by country (Adams 2007).

Physical Capital Accumulation and Foreign Direct Investment Most development models emphasize the role played by investment and physical capital accumulation in economic growth. For example, by raising capital per worker and productivity, the Solow model links increased savings and investment rates to higher income per capita.

Evidence on the role played by physical capital accumulation in economic growth varies by country, but it can be substantial. Young (1995) finds physical capital accumulation explains most of the miraculous economic growth of Singapore between 1966 and 1990. Recent work on China shows physical capital has also been key to its rapid growth since 1980 (see Riedel, Jin, and Gao 2007).

Most capital accumulation is financed through increased national savings. But in some countries, foreign capital contributes as well. Foreign direct investment (FDI) represents investments made by foreign residents (usually foreign firms) involving the acquisition of a lasting and significant control or ownership of the enterprise in which the funds are invested. Most FDI increases physical capital in the recipient countries and can stimulate economic growth. But until the early 1990s, FDI occurred mostly among high-income countries. In 1990, for example, developing countries accounted for only 12.1 percent of net FDI flows. This situation led the Nobel Prize winner Robert Lucas to write a paper in 1990 titled "Why Doesn't Capital Flow from Rich to Poor Countries?" The answer pointed mostly to restrictions and taxes imposed by developing countries on FDI inflows. This changed in the 1990s, with a massive wave of economic reforms that liberalized the flow of capital to developing countries. As a result, flows of FDI going to developing countries rose from $36.7 billion in 1990 to $314.6 billion in 2006 (measured in 2005 constant U.S. dollars).

Inflows of FDI have a direct, positive impact on employment. They can also accelerate the adoption of new technologies. If the FDI sector has significant

forward and backward linkages with other industries in the economy, positive productivity spillovers will develop, further stimulating growth (Rivera-Batiz and Rivera-Batiz 1990). Forward linkages may emerge from other industries that use the products sold by the foreign firms (such as domestic pharmaceuticals using chemicals produced by foreign petrochemical companies) while backward linkages may be in the form of domestic suppliers of business services or other inputs used by the foreign firms in their productive activities (such as a multinational's use of local transportation services). The case of Ireland, the so-called Celtic Tiger, has been widely discussed as an example of the positive effects of FDI on development. As much as 50 percent of employment in Irish manufacturing is in foreign-owned firms and some simulations suggest that Irish growth would have been two percentage points slower each year since 1990 without the FDI inflows (Navaretti and Venables 2004). But the empirical evidence on the positive effects of FDI is mixed. There are potentially negative indirect effects, including the displacement of local firms, environmental damage, and the deterioration of working conditions, as present in sweatshops or in the employment of child labor, although the latter often occur due to lax domestic regulation and may apply as well to domestic firms (see Nunnenkamp 2004; and Moran, Graham, and Blomstrom 2005).

Technological Change, Human Capital, and Brain Drain William Easterly and Ross Levine have argued that, among developing countries, capital accumulation does not explain most economic growth (Easterly and Levine 2001). Indeed, the presence of diminishing returns in the accumulation of capital suggests that long-run economic development prospects must be aided by forces other than physical capital investment. Technological change is one such force.

But what factors foster technological change? Building on the work of Paul Romer, Robert Lucas, Philippe Aghion, and Peter Howitt, among others, on endogenous growth theory, research on this issue has multiplied since the early 1990s (see Aghion and Howitt 1997). Most of this research suggests that modern technological change is spurred by new product creation and imitation. Countries with high rates of research and development expenditures are able to create new products that allow   albeit temporarily   rents that are reflected in higher growth rates. Imitators destroy those rents but at the same time provide incentives for new products and niches that generate further growth. This process, called "creative destruction" by Joseph Schumpeter, is postulated by this literature as a major determinant of development (Helpman 2004).

In the development literature, some researchers initially developed so-called product life-cycle models that visualized developing countries as imitators, reproducing products originally created in high-income countries. More recently, however, research based on the cases of South Korea, China, and other countries indicates that, by expanding sectors of the economy that are technologically dynamic, developing countries can bypass product cycles and spur innovation, providing internationally competitive new products that can then act as an engine of exports and growth. In this context, the process of development is stimulated not by how much you export but what you export (see Rodrik and Hausmann 2006).

The work of Paul Romer focused on how the domestic supply of human capital dedicated to research and development is the key factor influencing technical change (Romer 1990). Empirical evidence confirms that increased human capital can have a major stimulating impact on growth. This was the case in the so-called East Asian miracle economies Hong Kong, South Korea, Taiwan, and Singapore where increased quantity and quality of schooling accompanied their development process (Hanushek and Kimko 2000). But the evidence also shows that increased schooling does not universally have a positive impact on development (Pritchett 2001). In countries where the quality of the public sector governance is low, many highly educated workers tend to end up employed in low-productivity sectors (Murphy, Shleifer, and Vishny 1991). In some countries, increased school enrollments have led to a deterioration of quality. In these cases, investments

in human capital have a smaller impact on economic growth.

Domestic supplies of human capital can be severely depleted by emigration. The majority of migrants from developing countries are relatively unskilled but for some countries the emigration of professional and technical workers the so-called brain drain is most significant. The highest skilled emigration rates in the world are in the Caribbean, where as much as 43 percent of the region's labor force that has completed some tertiary (college) education currently resides outside its borders. There are also relatively high rates of skilled emigration in Africa, where the highest skilled-worker emigration rates include Cape Verde (68 percent) and Gambia (63 percent).

The emigration of skilled workers may not have a significantly negative impact on domestic consumers if it only causes a reduction in the output of sectors that produce tradable goods because these products can be imported from abroad when the laborers leave the country. But when the workers are employed in service sectors that produce internationally non-traded goods, the impact of emigration can be potentially disastrous because domestic consumers can only obtain those goods and services locally. If doctors and nurses emigrate, the supply of health services can collapse, resulting in higher prices and acute shortages. A brain drain can therefore reduce sharply the economic welfare of those left behind. Remittances may or may not offset these negative effects (Rivera-Batiz 2008).

If human capital flees an economy, then the ability of those left behind to sustain innovation may be compromised, thus reducing an economy's economic growth. On the other hand, the emigrants may generate international networks that enhance the scientific and technological capacities at home. Evidence on the relative significance of these two effects is not available.

Land, Agriculture, and Geography Geography can be a major barrier to development in several ways. First is the burden generated by diseases, which impose their cost more heavily in the tropics. Malaria, for example, is a disease that kills about 3 million people each year in tropical areas. It also weakens those who survive the disease. Countries that suffer from severe malaria epidemics grow more slowly than others and suffer from increased poverty (Gallup and Sachs 1999).

A second mechanism through which geography affects development is through transportation costs. Transportation costs act as a barrier to trade and, therefore, to development. Coastal economies or those close to navigable rivers generally have higher income per capita than landlocked economies without access to navigable waters. Indeed, outside of Europe, there is not a single high-income landlocked country. Much of the population in developing countries particularly in Africa lives in landlocked locations.

A third way geography influences development is through its impact on the agricultural sector. Tropical areas tend to suffer from climate influences that reduce agricultural productivity between 30 and 50 percent compared to temperate zones, holding other inputs constant (Gallup and Sachs 1999).

Public policies can be used to compensate for the impact of geography, whether through investments in health, transportation infrastructure, or irrigation schemes that can raise agricultural productivity. Land reform and rural market development policies are also essential for the development of agriculture. Despite the relatively scant attention paid to them by the World Bank and other international organizations, such policies have strong poverty reduction effects since rural areas are where the poor are concentrated. In the case of China, one of the most dramatic drops in poverty occurred in the early 1980s, before international trade became a major force in that country's development efforts. Rural poverty fell from 76 percent in 1980 to 23 percent in 1985, largely in response to comprehensive land reform initiatives and the growth of local agricultural markets. In the international arena, the Doha round of WTO negotiations begun in 2001 seeks to reduce agricultural subsidies and the non-tariff barriers to trade imposed by high-income countries, which hurt the exports of poor developing nations.

Governance and Institutions Although physical and human capital accumulation, technical change, demographics, land productivity, and so on are all proximate influences on development, they are all subject to the social, cultural, and economic environment within which individuals and firms live and function. This environment is closely connected to the country's government, which sets its legal and economic frameworks. Indeed, the quality of governance in a country is a major underlying determinant of development. High-quality governance includes, among other things, bureaucratic effectiveness and agility, institutions that assure the swift and strong accountability of government administrations, and a strong rule of law and control of corruption.

Since the mid-1990s, the World Bank has been collecting data on the public sector governance of most countries in the world. Based on these data, the evidence shows that good public sector governance is essential for development (Kaufmann, Kraay, and Mastruzzi 2007). A case in point is the acceleration of economic growth in India, which can be seen as the outcome of governance reforms in the 1980s (Rodrik and Subramanian 2005). Nations able to generate high quality of public sector governance   such as some of the East Asian miracle economies, including Singapore and Hong Kong   have been able to cultivate institutions that complement the process of economic development, introducing policies that foster higher investment in physical and human capital and provide incentives for entrepreneurship and innovation.

But what forces generate poor or good country governance? A major force is political power. Different interest groups or classes within a country have conflicting interests over the distribution of resources, and their policy prescriptions are informed by such interests. Political institutions then determine how various groups get   or do not get   access to government and its legislative powers. Overall, the quality of governance is substantially higher in more democratic countries, even after holding other variables constant. Transparent democratic institutions mean that, over the long run, inept and corrupt officials will be voted out of office,

which is often more difficult under authoritarian regimes. More democratic institutions also facilitate the activities of the press and allow the dissemination of information to the public so that governments can be held accountable (Rivera-Batiz 2002). At the same time, there are cases of authoritarian regimes   such as Singapore and China   that have been able to provide high-quality public sector governance, spurring economic growth.

In many developing nations, a history of colonialism provides a backdrop for the creation of institutions that can hinder development efforts after independence (Acemoglu et al. 2006). Social conflict can also break down the government stability necessary for stable growth. Such conflict is endemic in the poorest developing countries and is fostered by relatively high proportions of poor young men with little education, ethnic or religious strife, and the presence of natural resources (such as diamonds or oil), which finance the armed conflicts (Collier 2007). Once social conflict starts, it generates a vicious circle, where conflict leads to poverty and poverty to more conflict. The more fragmented a country is   by ethnicity or social class   the lower its rate of economic growth. Many countries in Africa appear to follow this pattern (Easterly and Levine 1997).

Natural Resources and the Dutch Disease The presence of large rents can also give rise to poor governance and, as a result, constrain economic growth. Consider the case of countries that have rich endowments of natural resources. Surprisingly, the evidence shows that, despite their potential wealth, these countries have not grown faster nor have they been able to reduce poverty rates, holding other things constant (Sachs and Warner 1995b).

Consider the case of Nigeria, where massive oil reserves were discovered in 1965. The sum of oil revenues received by this country since that time has been more than $400 billion but its gross domestic product (GDP) per capita was about the same in 2000 as it was in 1970, when adjusted for inflation. In the meantime, poverty rates (using a $1 a day measure) rose from 36 percent in 1970 to 70 percent in 2000.

The lack of impact of natural resource exports on growth and poverty may be connected to the fact that, as countries pull resources into the exploitation of natural resources, they withdraw resources from other sectors. As a consequence, there is a crowding-out effect, with natural resource output crowding out manufacturing output, leaving no net impact on GDP. This has been called *Dutch disease*, referring to the case of natural gas exploitation in the Netherlands in the 1970s, which failed to generate sustained economic growth.

A second explanation links the increased exploitation of natural resources to a lower rate of technical change. The rewards for working in the natural resource industry are high and therefore a significant portion of a country's talent may end up employed in this industry. But the natural resources sector is not itself one that leads to great innovations or stimulates entrepreneurship. As a result, countries specializing in the production of natural resources may in fact have low rates of innovation and face lower long-run growth.

But in the case of Nigeria and many other countries, the culprit behind the lack of impact of natural resource exports on economic growth is poor public sector governance (Sala-i-Martin and Subramanian 2003). The consequence of massive economic rents is the emergence of a governance regime that permits those with greater political power to capture, by any means necessary, some of those rents. The result is corruption and slow growth.

The impact of natural resources on a country can be transformed from a curse to a blessing. For instance, some nations have been able to ameliorate the rent-seeking and corruption that is often associated with the export of natural resources. Part of the strategy has been to preempt possible corruption by earmarking or assigning a share of natural resource revenues directly to finance social programs, such as education and health. In the case of Indonesia, for example, revenues from oil exports were directed to a massive program of investment in education. More recently, Chile imposed a tax on copper production in order to finance a Competitiveness Innovation Fund that finances new technology projects. Through the use of innovative government policies, the exploitation of natural resources can have positive effects on development (Stiglitz 2005).

International Trade and Commercial Policy
The impact of international trade on economic development has been at the center of the debate on the consequences of globalization and is one of the most controversial issues in development economics. The key questions are whether trade liberalization reduces or increases poverty and inequality and whether it has a positive impact on economic growth.

The HOS model postulates that increased trade should reduce both poverty and inequality in developing countries. Trade induces a developing country to increase the production of goods in which it has a comparative advantage, that is, in sectors that use unskilled labor intensively. As production and export of these products rises, the demand for unskilled workers increases and this raises their relative wages. At the same time, the demand for and the rewards received by human and physical capital decline. Since unskilled workers are usually poor while the owners of both physical and human capital tend to be richer, the impact of international trade in the HOS framework is to reduce poverty and improve income distribution in developing countries.

The evidence on the impact of trade on poverty tends to support the HOS predictions. The data on extreme poverty, measured as people with consumption below the $1 a day level, show that world poverty has dropped sharply during the period 1981 to 2004, when globalization bloomed. Poverty declined from 40.1 percent in 1981 to 18.1 percent in 2004. This represents a halving of poverty rates and has meant a reduction in the number of poor in the world from 1,470,000 in 1981 to 969,480 in 2004 (Chen and Ravallion 2007). Furthermore, the two economies that have seen some of the sharpest increases in trade, China and India, are also the two economies where poverty has dropped the most. To cap off all of this, the region that has been the slowest to drop trade barriers, sub-Saharan Africa, is also the region where poverty failed to drop during the period.

On the other hand, the data show much less impressive drops in poverty when the poverty line rises to $2 a day. And metastudies that include a wide variety of developing countries show cases where poverty particularly regional poverty has risen as a result of trade (Harrison 2006). Furthermore, most of the estimates available suggest that the recent expansion of international trade in the world has been associated with a period of increased inequality. This inequality is displayed in both greater within-country inequality and higher cross-country inequality (Milanovic 2005).

The increasing inequality and declining poverty associated with globalization can be reconciled by the widespread evidence indicating that trade is linked to greater economic growth (Dollar and Kraay 2004; Wacziarg and Welch 2004). Economic growth tends to raise all boats and is associated with a reduction of poverty. But it has also been connected to rising inequality. One reason is that long-run growth in recent years has been closely linked to technological change. And this technological change has tended to be skill-biased, raising the demand for skilled labor at the expense of unskilled labor. Such a shift in demand raises the wages of skilled labor relative to unskilled workers, increasing inequality (Rivera-Batiz 2007).

For increased international trade to stimulate economic growth, domestic investment rates must be sufficiently high that new export industries can be developed. The evidence available suggests that, from South Korea to China, the implementation of active government policies designed to stimulate investment have been essential in allowing trade liberalization to generate growth and reduce poverty (Rodrik 1995). Without these and other complementary forces the impact of trade on growth is limited.

Financial Development and International Financial Markets One of the major forces that serves as a backdrop for economic growth is financial development. Although there is debate about the role of specific financial institutions, a growing body of empirical work shows that a well-functioning financial system has a strong, positive impact on economic growth and helps alleviate poverty by allowing the efficient transformation of savings into productive investments (Beck, Demirgüç-Kunt, and Levine 2007). The financial sector facilitates the trading, diversification, and management of risk, and it helps mobilize and pool savings by allowing savers to hold liquid assets and transforming those into higher-return long-term investments (Levine 2005).

While providing a solid, long-run basis for overall economic growth, the financial sector has nonetheless been prone to crises, with disastrous short-term consequences. In addition, traditional financial institutions have failed to fund the poor, who lack the collateral to access credit. In recent years, however, myriad microfinance institutions have emerged to provide credit to the poor. Microfinance institutions minimize the impact of a lack of collateral by pooling responsibility for loan repayment among communities of borrowers (Armendariz and Murdoch 2005). In addition, by having mobile operators who move among villages, they minimize transaction costs and mitigate adverse selection (which occurs when lenders cannot monitor the riskiness attached to borrowers' projects).

The conventional wisdom is that, as countries develop, the limits imposed by domestic savings may constrain development. From this perspective, international financial liberalization is a good thing for development, allowing developing countries with scarce capital to borrow in order to finance greater investment without the necessity of raising domestic savings. In addition, financial capital flows permit residents of different countries to pool risks, achieving more effective insurance than purely local arrangements would allow (Obstfeld and Taylor 2005; Prasad et al. 2003). International financial liberalization can also increase the efficiency of the domestic financial system, whether by enhancing the liquidity of local stock markets or by promoting contested markets and raising the efficiency of domestic banking.

But international capital market liberalization has its critics. Some argue that there is no systematic evidence that financial capital inflows cause greater

growth. Others observe that unregulated capital inflows can be disastrous for developing countries, pointing to the rash of financial crises in Mexico, East Asia, Russia, Argentina, and a variety of other emerging markets during the 1990s and early 2000s (Stiglitz 2003). These crises occurred after capital flows to developing nations began to rise sharply in the late 1980s and early 1990s in response to the elimination of many restrictions on international capital movements.

There were very specific factors connected to the East Asian and other emerging market crises of the 1990s and early 2000s. Fragile domestic financial institutions were unable to handle efficiently the expansion of credit, partly because they lacked the appropriate expertise and partly because governments had failed to implement adequate regulatory and supervision systems. The result was overinvestment, in the form of excessive lending to risky projects with low short-term returns. In addition, the reliance on highly volatile, short-term debt led to liquidity crises that resulted in financial panics and bank runs. These problems were compounded by the fact that the capital inflows caused real currency appreciation by generating inflation in fixed exchange-rate regimes. As the value of the currency rose in real terms, the current account balance deteriorated and external borrowing climbed. These changes then led to expected currency depreciations that fueled capital flight. When central banks stubbornly defended prevailing currency values, their foreign exchange reserves were eventually depleted, precipitating a foreign exchange crisis that spilled over into a financial crisis as well (Rivera-Batiz 2001).

Most countries in financial crisis have benefited from massive loan packages provided by the International Monetary Fund (IMF). But the conditions imposed by the fund to disburse its loans were sometimes so stringent that they sharply increased poverty in the short run and may have undermined future growth. The next section examines the role in development played by international organizations.

Foreign Aid and Development Many of the poorest developing countries lack access to world private capital markets. For these countries, official development assistance (ODA) grants and concessional loans made by governments and international organizations constitutes a major source of capital. In 2005, a total of $120.4 billion of ODA flowed to developing nations. This foreign aid can originate in other, mostly high-income countries (bilateral aid) or in groups of countries (multilateral aid), including international organizations such as the United Nations, the World Bank, and the IMF.

The impact of foreign aid on development has been a source of controversy, with proponents of aid (such as Sachs 2005) arguing for much greater flows while critics (such as Easterly 2006) argue that such aid is largely a waste of resources. Evidence of the positive impact of specific foreign aid projects on development is widely available. International efforts that funded campaigns to rid developing countries from various diseases, including smallpox, polio, and guinea worm disease, among others, have been found to be successful (Levine 2007). The role played by country donors and international organizations in the development of new seed strains as well as technological innovations now referred to as the Green Revolution is well known. And recent studies have used random experimental designs to show the success of educational and other development projects (such as the Progresa/Oportunidades program in Mexico), projects which have received substantial foreign aid support.

But many foreign aid projects have failed miserably. For instance, World Bank educational projects in the 1980s and early 1990s often requested that countries introduce public school fees in order to raise additional funding. But the outcome of those policies was to sharply reduce school enrollment. Another example is the possibility that the various market-oriented policies adopted by the World Bank and IMF toward agricultural markets may have undermined rural development efforts in low-income countries, especially in sub-Saharan Africa. Even the World Bank's own operations evaluation department finds that in the period 1995 to 2004, at least one-third of all projects in sub-Saharan Africa failed

to achieve their goals, more than half were unlikely to be sustained over time, and the great majority of projects did not have a substantial institutional impact on developing country governments. Some of this failure can be attributed to poor country governance. Nonetheless, systematically positive effects of foreign aid on overall economic growth have not been found (Rajan and Subramanian 2005).

A similar lack of support exists for the lending policies of the IMF. Despite the presence of apparently successful cases, such as Chile in the 1980s, South Korea in the 1990s, and Brazil in the 2000s, in other countries IMF-sponsored short-term reforms have not fared as well, including Bolivia in the 1980s, Indonesia in the 1990s, and Argentina in the 2000s. The evidence on the long-run growth effects of lending provided by the IMF under its conditionality agreements is mixed, with some studies finding negative impacts, especially among low-income countries (Vreeland 2003).

Macroeconomic Policies Notwithstanding the issue of the effectiveness of IMF stabilization policies, the IMF does not provide access to its lending facility unless asked by member countries in economic or financial distress. The available research suggests that macroeconomic mismanagement can be a major barrier to development (Montiel 2002).

Although the evidence on the relationship between low or moderate inflation and growth is mixed, the overwhelming evidence is that high inflation reduces growth and raises poverty (Barro 1995). By undermining the financial system and generating uncertainty, high and variable inflation rates inhibit investment and hurt growth. By acting as a tax on money balances, affecting especially those with fixed incomes, high inflation also hurts the poor.

Central banks, which manage the money supply, usually have as one of their goals the control of inflation. But a phenomenon economists refer to as time inconsistency means that, insofar as central banks are under the influence of the rest of the government, they may have an inflationary bias. This is associated with the fact that central banks have an incentive to break their own commitments to maintain low inflation rates because of the political

benefits of engaging in expansionary policies that provide greater output and employment in the short run (at the expense of higher inflation). As a result, countries that have more independent central banks also tend to have lower inflation (Cukierman 1992).

Government budget deficits can also act to constrain economic growth (Adam and Bevan 2005). Once the deficits rise over a certain level, the growing government financing needs can give rise to a wide array of negative consequences. Private investment may be crowded out by higher interest rates. And if the deficit is financed by borrowing, a ballooning external debt can raise country risk, leading to an expected currency depreciation and capital flight. If monetized, rising budget deficits can also lead to high inflation.

Currency regimes can provide a major boost to development or serve as a constraint. Both fixed and flexible exchange rate regimes have pros and cons (Agenor and Montiel 1999, chapter 7). Fixed exchange rates facilitate a stable trading environment that can foster exports and imports of goods and services. Therefore, among countries that are closely integrated, a fixed exchange rate regime or maybe even a common currency may be the optimal regime. But among countries that are not closely integrated, with substantial net flows of international capital, fixed exchange rate regimes become potentially subject to speculative attacks and to currency crises. Flexible exchange rates may allow the more stable management of capital flows and can serve as a policy instrument for the adjustment of relative prices. Furthermore, monetary policy loses its effectiveness in a regime of perfect capital mobility and fixed exchange rates. A mixture of fixed and flexible exchange rates may be the optimal outcome for some countries.

The variety of experiences with development policies reflects the complexity of development itself. As shown earlier, the links of development to the world economy are also complex and bring challenges as well as opportunities.

See also aid, international; capital accumulation in open economies; development; evolution of development thinking; foreign direct investment (FDI); global income

inequality; migration, international; poverty, global; trade and economic development, international

FURTHER READING

Acemoglu, Daron, Simon Johnson, and James Robinson. 2006. "Understanding Prosperity and Poverty: Geog raphy, Institutions, and the Reversal of Fortune." In *Understanding Poverty*, edited by A. V. Banerjee, R. Benabou, and D. Mookherjee. New York: Oxford University Press, 19 35. A survey of the historical and political factors that influence institutions and public sector governance in a country. The authors make the case that these factors explain more directly differences in poverty and prosperity around the world than geog raphy or other determinants of prosperity.

Adam, Christopher S., and David L. Bevan. 2005. "Fiscal Deficits and Growth in Developing Countries." *Journal of Public Economics* 89 (3): 571 97. Examines the issue of whether government budget deficits affect a coun try's economic growth. Evidence is presented showing that fiscal deficits are indeed negatively related to growth in developing countries.

Adams, Richard. 2007. "International Remittances and the Household: Analysis and Review of Global Evidence." Policy Research Working Paper No. 4116. Washington, DC: The World Bank. Provides a survey of the litera ture on the determinants and consequences of mi grant remittances in the world, including the role of these remittances in the economic welfare of the migrants, their families, and the countries involved.

Agenor, Pierre Richard, and Peter Montiel. 1999. *Devel opment Macroeconomics*. Princeton, NJ: Princeton Uni versity Press. A textbook on the theory and evidence of how macroeconomic factors affect development. It goes over a variety of theoretical models, describing them and providing evidence on a wide range of issues, including the effects of fiscal and monetary policies, credibility and disinflation, exchange rate regimes, currency substitu tion, speculative attacks, and political economy, among many others.

Aghion, Philippe, and Peter Howitt. 1997. *Endogenous Growth Theory*. Cambridge, MA: Harvard University Press. This book offers a technical review of the literature on endogenous growth, focusing on the various deter minants of innovation and technological change and how research and development, education, competition, trade, policy, and other forces affect growth.

Armendariz, Beatriz, and Jonathan Murdoch. 2005. *The Economics of Microfinance*. Cambridge, MA: MIT Press. Presents a comprehensive survey of the theory of how microfinance institutions work and the empirical evi dence on their worldwide growth and effectiveness. It examines the market failures and economic phenom ena that plague lending for the poor including capi tal market imperfections, moral hazard, and adverse selection and the mechanisms used by microlending institutions to deal with them.

Barro, Robert. 1995. "Inflation and Economic Growth." NBER Working Paper No. 5326. Cambridge, MA: National Bureau of Economic Research. This paper presents research on how inflation is linked to growth. It includes a discussion of the various forces involved as well as empirical evidence from a cross section of countries.

Beck, Thorsten, Asli Demirgüç Kunt, and Ross Levine. 2007. "Finance, Inequality, and the Poor." *Journal of Economic Growth* 12 (1): 27 49. Presents the existing research on the links between financial development, growth, and poverty. The authors argue that, despite popular beliefs, there are a number of mechanisms through which financial development acts to reduce poverty and inequality.

Bhagwati, Jagdish N. 2004. *In Defense of Globalization*. Princeton, NJ: Princeton University Press. A wide ranging review of the arguments in favor of and against globalization. The author argues that, overall, global ization has had a positive influence on both economic growth and the standards of living of the poor.

Bloom, David E., and Jeffrey Williamson. 1998. "Demo graphic Transitions and Economic Miracles in Emer ging Asia." *World Bank Economic Review* 12 (3): 419 55. This paper presents research on how demo graphic forces affect economic development. The au thors make the argument that a demographic dividend explains a significant part of the economic growth ex plosion in East Asia.

Chen, Shaohua, and Martin Ravallion. 2007. "Absolute Poverty Measures for the Developing World, 1981 2004." Development Research Group Working Paper No. 4211. Washington, DC: The World Bank. Presents

the most recent World Bank estimates on poverty in the world and its behavior over time. The paper shows that absolute poverty has generally declined in the developing world since 1981, especially in certain countries and regions, such as East Asia. It also shows that specific countries and regions such as sub Saharan Africa   have failed to display any significant reductions in poverty.

Collier, Paul. 2007. *The Bottom Billion: Why the Poorest Countries Are Failing and What Can be Done About It.* Oxford: Oxford University Press. Arguing that poverty has dropped sharply in most of the developing world except for about 50 countries ("the bottom billion"), this book seeks to determine why these countries have been unable to prosper. The author discusses a set of traps that prevent economic development and shows how traditional policies including foreign aid and globalization   have failed to improve matters.

Cukierman, Alex. 1992. *Central Bank Strategy, Credibility, and Independence: Theory and Evidence.* Cambridge, MA: MIT Press. A technical survey of the theory and evidence on how central banks function and how they affect the macroeconomy. Foremost among the various issues discussed is how central bank independence has been linked to lower inflation.

Dollar, David, and Aart Kraay. 2004. "Trade, Growth, and Poverty." *Economic Journal* 114 (2): F22 F49. This paper presents research showing that international trade reduces poverty by stimulating economic growth.

Easterly, William. 2006. *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good.* New York: Penguin. A wide ranging study of how, after decades of financial aid valued at over $2.3 trillion, the development strategies forged in the West have failed to produce sustained development. The author argues that development efforts should refocus on specific microprojects in areas such as education and health, which have been found to be successful in reducing poverty.

Easterly, William, and Ross Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *Quarterly Journal of Economics* 112 (3): 1203 50. Dcuments the substantial racial and ethnic fractionalization in African countries and the negative impact such divisions have had on economic growth in the region.

. 2001. "It's Not Factor Accumulation: Stylized Facts and Growth Models." *The World Bank Economic Review* 15 (2): 177 219. Presents research on the relative roles played by physical and human capital accumulation as well as technological change in explaining differences in growth among a cross section of countries. The authors show that technological change appears to be the dominant factor.

Gallup, John Luke, and Jeffrey D. Sachs. 1999. "Geography and Economic Development." Center for International Development Working Paper No. 1, Cambridge, MA: Harvard University. Presents theoretical arguments and empirical evidence on how health, transportation costs, and agricultural productivity are affected by geography.

Hanushek, Eric A., and Dennis D. Kimko. 2000. "Schooling, Labor Force Quality, and the Growth of Nations." *American Economic Review* 90 (5): 1184 1208. Using the results of international assessments of education to measure the differences among various countries, this article offers a technical analysis of how the quantity and quality of schooling influence economic growth.

Harrison, Anne. 2006. *Globalization and Poverty.* Chicago, IL: University of Chicago Press. This volume provides a comprehensive survey of the various links between globalization, poverty, and inequality in the world. It shows that there exist diverse country experiences of the impact of rising trade on the poor and how it has affected income inequality.

Helpman, Elhanan. 2004. *The Mystery of Economic Growth.* Harvard, MA: Harvard University Press. This book offers a survey of the determinants of economic growth, including the roles played by physical and human capital, productivity growth, technological change and its determinants, interdependence across countries, inequality, and institutions and policies.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2007. "Governance Matters VI: Aggregate and Individual Governance Indicators for 1996 2006." Policy Research Working Paper No. 4280. Washington, DC: The World Bank. This paper presents the latest World Bank indicators on public sector governance in the world. It also shows the connections between improved governance and various indicators of economic development.

Levine, Ross. 2005. "Finance and Growth: Theory and Evidence." In *Handbook of Economic Growth*, edited by

Philippe Aghion and Steven Durlauf. The Nether lands: Elsevier Science, 865 934. A survey of how financial development affects economic growth. The author makes a strong case that policies that promote financial development have a significant positive impact on growth.

Levine, Ruth. 2007. *Case Studies in Global Health: Millions Saved*. Boston: Jones and Bartlett. This volume presents examples of successful foreign aid and domestic policy interventions in the field of global health in the devel oping world. It shows how international development assistance has been effective in some cases and draws lessons on how such assistance can work more effectively in other contexts.

Mankiw, N. Gregory, David Romer, and David N. Weil. 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107 (2): 407 37. This paper presents empirical evidence showing how population growth and the accumulation of phys ical and human capital relate to income per capita and economic growth.

Milanovic, Branko. 2005. *Worlds Apart: Measuring Inter national and Global Inequality*. Princeton, NJ: Princeton University Press. This book presents evidence on the disparities between poor and rich countries. Examining various measures of inequality, the author utilizes data from household surveys from over 100 countries to study the extent of global inequality and how it has changed over time.

Mishra, Prachi. 2006. "Emigration and Wages in Source Countries: Evidence from Mexico." Working Paper No. WP/86/06. Washington, DC: International Monetary Fund. Using data from 1970 2000, this paper examines the impact of emigration on the labor market and economy of Mexico. It shows that emigration has re sulted in a rise of wages for those who remain in Mexico, including those with high levels of human capital.

Montiel, Peter. 2002. *Macroeconomics in Emerging Markets*. Cambridge, MA: Cambridge University Press. This volume provides an integrated, technical discussion of the main macroeconomic issues facing developing countries. It presents both theory and evidence, includ ing discussions of high inflation episodes, central bank ing, financial repression, banking crises, and exchange rate management.

Moran, Theodore H., Edward M. Graham, and Magnus Blomstrom, eds. 2005. *Does Foreign Direct Investment Promote Development?* Washington, DC: Institute for International Economics. The authors present research on the diverse experiences that developing countries have had with the consequences of foreign direct investment. Ranging from Turkey to East Africa, they present a num ber of case studies, including an analysis of productivity spillovers to domestic firms and the impact on research and development and on country economic growth.

Murphy, Kevin M., Andrei Shleifer, and Robert Vishny. 1991. "The Allocation of Talent: Implications for Growth." *Quarterly Journal of Economics* 106 (2): 503 30. This paper presents research on how the sectoral allocation of educated and professional workers in de veloping countries can have positive or negative impacts on development. For example, if the talented persons in a country end up employed in profitable but static low growth sectors due to poor governance and rent seeking activities, then increased education may have a limited impact on development.

Navaretti, Giorgio Barba, and Anthony J. Venables. 2004. *Multinational Firms in the World Economy*. Princeton, NJ: Princeton University Press. A comprehensive survey of multinational firms, this book examines why multi nationals exist, what makes them effective, their impact on both host countries and source countries, and their policy implications.

Nunnenkamp, Peter. 2004. "To What Extent Can Foreign Direct Investment Help Achieve International Devel opment Goals?" *World Economy* 27 (5): 657 77. This paper examines the potential effects of foreign direct investment in developing countries. It shows that, de spite high expectations, the impact of foreign direct in vestment varies by country and depends on the charac teristics of host countries and their policies toward foreign direct investment.

Obstfeld, Maurice, and Alan Taylor. 2005. *Global Capital Markets: Integration, Crisis, and Growth*. Cambridge, MA: Cambridge University Press. Provides a historical view of international capital flows in the world. Showing that the recent globalization of capital flows constitutes a reversal to an earlier period of great capital mobility in 1880 1914, it examines the differences between these two golden periods for capital flows.

Prasad, Eswar, Kenneth Rogoff, Shang Jin Wei, and M. Ayhan Kose. 2003. "Effects of Financial Globaliza tion on Developing Countries: Some Empirical Evi dence." IMF Occasional Paper No. 220. Washington, D.C.: International Monetary Fund. This report ex amines the impact of the liberalization of international financial flows on the economic growth and consump tion volatility faced by developing countries. Both the oretically and empirically, the authors find that financial liberalization has mixed effects. The overall impact on growth may be tenuous and the impact on volatility could be negative for some countries. They suggest that more positive effects of international financial liber alization may be obtained by countries that have the adequate institutions and domestic absorptive capacity necessary to handle the liberalization.

Pritchett, Lant. 2001. "Where Has All the Education Gone?" *The World Bank Economic Review* 15 (3): 367 92. This paper presents research showing that in creased educational attainment may not be connected to greater economic growth. Empirical evidence is pre sented along with possible explanations, including pos sible data problems, rent seeking by educated workers, and issues with the quality of the increased schooling.

Rajan, Raghuram G., and Arvind Subramanian. 2005. "Aid and Growth: What Does the Cross Country Evidence Really Show?" Working Paper No. 11513. Cambridge, MA: National Bureau of Economic Research. This ar ticle examines the empirical links between foreign aid and economic growth using panel data for a cross section of countries. The authors fail to find any evidence that foreign aid has a positive effect on economic growth.

Riedel, James, Jing Jin, and Jian Gao. 2007. *How China Grows: Investment, Finance, and Reform*. Princeton, NJ: Princeton University Press. This volume examines the various factors that have been associated with the recent rise of economic growth in China. From increased sav ings and investment to policies toward technological change, the book argues that the Chinese economic miracle is related to a wide array of economic forces.

Rivera Batiz, Francisco L. 1983. "Trade Theory, Distribu tion of Income, and Immigration." *American Economic Review* 73 (2): 183 87. This paper examines the theo retical arguments of how migration affects the distribu tion of income. It concludes that although in the short run immigration can reduce wages in the host country and emigration can raise wages in the source country, there are long run forces that reverse these effects. As a result, the long run impact of migration on wages in host and source countries may be minimal.

. 2001. "The East Asian Crisis and the Anatomy of Emerging Market Disease." In *The Political Economy of the East Asian Crisis: Tigers in Distress*, edited by Arvid Lukauskas and Francisco L. Rivera Batiz. Cheltenham, UK: Edward Elgar, 31 73. Examining the causes and consequences of the East Asian financial crisis, the au thors argue that the crisis was precipitated by the impact of massive, unrestricted capital inflows in the late 1980s and the 1990s. These inflows generated inflation, a de terioration of domestic competitiveness in interna tional markets, current account balance deficits, and increased indebtedness, all of which eventually led to expectations of devaluation and to the massive exodus of the foreign capital, precipitating exchange rate and financial crises.

. 2002. "Democracy, Governance and Economic Growth: Theory and Evidence." *Review of Development Economics* 6 (2): 225 47. This article shows the theory and evidence on how democracy is related to growth. The conclusion is that the most important determinant of economic growth is the quality of the public sector governance in a country, independent of its political rights situation. However, evidence is also presented that, in the long run, having a larger number of demo cratic institutions is connected to greater quality of governance.

. 2007. "Trade, Poverty, Inequality, and Gender." Trade Policy Executive Program. Washington, D.C.: The World Bank. A survey of the literature on the im pact of international trade on poverty and inequality, including its effects on gender inequities. The article concludes that although globalization has been associ ated with a general drop in poverty, it has also resulted in increased inequality. Furthermore, trade liberaliza tion has had a negative impact on specific groups and sectors in developing countries. Its impact on gender inequities has been mixed and depends on the country and region being considered.

. 2008. "International Migration and the Brain Drain." In *International Handbook of Development Economics*, edited by Amitava Krishna Dutt and Jaime Ros. Cheltenham, U.K.: Edward Elgar. Presents evidence on the extent of skilled emigration from developing countries in recent years and its consequences for the migrants themselves as well as for the source country and host country populations.

Rivera Batiz, Francisco L., and Luis A. Rivera Batiz. 1990. "The Effects of Direct Foreign Investment in the Presence of Increasing Returns Due to Specialization." *Journal of Development Economics* 34 (1): 287 307. This paper shows the theoretical connections between direct foreign investment and host country economic welfare. It demonstrates how foreign direct investment can provide positive gains to the host country through its employment effects as well as the potentially positive spillovers associated with the increased demand for local producer services by the foreign companies.

Rodrik, Dani. 1995. "Getting Interventions Right: How South Korea and Taiwan Grew Rich." *Economic Policy* 20: 55 107. This article examines why South Korea and Taiwan have been able to generate rapid growth. The author concludes that although the popular explanation has relied on export promotion and trade liberalization, the main factor that led to the economic miracles in these countries was increased investment rates (in physical and human capital), spurred by a wide array of government policies and incentives.

Rodrik, Dani, and Ricardo Hausmann. 2006. "What's So Special about China's Exports?" Working Paper RWP06 001. John F. Kennedy School of Government. Cambridge, MA: Harvard University. Examines the extent to which China's exports consist of products that have the technological sophistication of goods usually produced by higher income countries. The authors find evidence that, despite the popular view of Chinese exports as cheap, low technology products, a significant portion does consist of goods with substantial technological sophistication.

Rodrik, Dani, and Arvind Subramanian. 2005. "From Hindu Growth to Productivity Surge: The Mystery of the Indian Growth Transition." *IMF Staff Papers* 52 (2): 193 228. Although the popular explanation for the Indian growth spurt has been its trade liberalization measures, the authors of this paper argue that before trade was liberalized, India engaged in institutional, public sector reforms that allowed trade liberalization, as well as other economic changes, to take place.

Romer, Paul. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5): S71 102. Presents a theoretical framework examining the determinants of technological change. The focus is on how human capital employed in research and development sectors is the key to technological change and economic growth.

Sachs, Jeffrey D. 2005. *The End of Poverty: Economic Possibilities for Our Time*. New York: Penguin Press. This volume makes the case for increased worldwide aid in the developing world. It argues that such aid can be effectively used to reduce poverty by investing it in health, education, infrastructure, and other projects.

Sachs, Jeffrey D., and Andrew Warner. 1995a. "Economic Reform and the Process of Global Integration." *Brookings Papers on Economic Activity* 1: 1 118. This article examines the links between increased international trade and economic growth. By developing a comprehensive index of openness to international trade, the authors are able to examine empirically the connection between such openness and economic growth in a cross section of countries. They find a strong, positive connection between openness and growth.

. 1995b. "Natural Resource Abundance and Economic Growth." NBER Working Paper No. 5398. Cambridge, MA: National Bureau of Economic Research. This paper studies the empirical connection between exports of natural resources and economic growth. Surprisingly, it finds that countries that have a greater share of natural resources in their exports tend to grow more slowly, holding other things constant.

Sala i Martin, Xavier, and Arvind Subramanian. 2003. "Addressing the Natural Resource Curse: An Illustration from Nigeria." Working Paper WP/03/139. Washington, DC: International Monetary Fund. This article examines why the oil exporting country of Nigeria has been unable to develop economically since oil reserves first began to be exploited decades ago. Although several hypotheses are explored, the authors conclude that the most significant factors were the low quality

public sector governance prevailing in the country and the increased corruption associated with the rent seeking connected to the oil exports.

Stiglitz, Joseph. 2003. *Globalization and its Discontents.* New York: W. W. Norton. This book offers a dissection of the various effects that globalization has had on the world economy. Drawing on his extensive experience as a researcher and policymaker, the author presents a wide array of criticisms of the so called Washington consensus, a set of market oriented policies adopted by the United States which have favored free trade, deregulation, privatization, etc. According to the author, these policies have often had disastrous results in the developing world.

. 2005. "Making Natural Resources into a Blessing Rather Than a Curse." In *Covering Oil*, edited by Svetlana Tsalikad and Anya Schiffrin. New York: Open Society Institute, 13 20. Presents a set of policies that can be used by countries with large endowments of natural resources to ensure that the revenues from the exploitation of such resources help rather that hurt the development process.

Stiglitz, Joseph E., and Andrew Charlton. 2005. "Trade Can Be Good for Development." In *Fair Trade for All: How Trade Can Promote Development*, edited by J. Stiglitz and A. Charlton. New York: Oxford University Press, 11 40. Shows how international trade can have positive or negative effects on development. It specifies the conditions under which trade can be harmful and those under which it can be more beneficial.

Vreeland, James R. 2003. *The IMF and Economic Development.* New York: Cambridge University Press. This book presents a wide ranging discussion of the impact of IMF stabilization policies on economic development. Providing an analysis that ranges from the quantitative to the qualitative and from economics to political science, the book offers a comprehensive yet critical view of the effect that these policies have had on developing countries.

Wacziarg, Romain, and Karen Horn Welch. 2004. "Trade Liberalization and Growth: New Evidence." Working Paper No. 7. Stanford, CA: Center on Development, Democracy and the Rule of Law, Stanford Institute for International Studies. This paper reports research on the links between trade liberalization and economic growth. By looking at the experience of a wide array of countries before and after trade liberalization, it shows that, generally, trade liberalization is associated with greater economic growth, although a number of exceptions are also discussed.

Young, Alwyn. 1995. "A Tale of Two Cities: Factor Accumulation and Technical Change in Hong Kong and Singapore." *NBER Macroeconomics Annual 1995.* Cambridge, MA: MIT Press. This paper examines empirically the determinants of economic growth in Hong Kong and Singapore. The author concludes that Singapore grew primarily because of its investments in physical capital, while Hong Kong's growth drew more strength from productivity gains connected to technological developments and innovation.

FRANCISCO L. RIVERA BATIZ

■ economic geography

See New Economic Geography

■ economies of scale

International trade theory recognizes three fundamental reasons for countries to trade: *comparative advantage* (to exploit differences in countries' tastes, technologies, or factor endowments), *economies of scale* (to concentrate on fewer tasks in order to produce more efficiently), and *imperfect competition* (to expose firms to more competition). Comparative advantage has always been dominant in trade theory, although economies of scale also long played a secondary role. This changed in the late 1970s, when economists realized that the lion's share of world trade consisted of the exchange of similar (manufactured) goods between similar (rich) countries.

A common reaction to this realization was that such trade could not be due to comparative advantage. But if similar countries trade similar goods, price elasticities (the sensitivities of demands and supplies of goods to price variations) are likely to be high since the traded products will likely be good

substitutes. Similar countries would have similar relative prices in autarky (the absence of international trade), so comparative-advantage trade, establishing a world price between the two autarky prices, would imply a modest price change in each country. But, with high elasticities, this could still involve heavy trade. Likewise, dissimilar countries trading very distinct goods   manufactures for primary products, for example   could be expected to experience large price changes, but the resulting trade volumes could be small as price elasticities are likely to be low.

So, comparative advantage was not necessarily inconsistent with actual trade. Nevertheless, it was important to consider alternative possibilities because, although trade to exploit differences does *not* imply that the greater the differences, the greater the trade, it does imply, other things being equal, that the greater the differences, the greater the *gains* from trade. Thus one might conclude that the smaller part of world trade   that between dissimilar countries in quite different products   is more important for policy. But if trade patterns are significantly due to something other than comparative advantage, this conclusion need not follow.

Increasing returns to scale (IRS) and imperfect competition supply alternatives. This article considers the former, often also referred to as economies of scale.

Types of Scale Economies In practice, scale economies occur in great variety, so a classification of the more important attributes is useful.

Internal versus external (to the firm). Scale economies are internal if the individual firm can reduce average costs by operating at a higher scale (e.g., assembly-line operations and equipment made possible by large-scale production). They are external if the individual firm operates subject to constant returns to scale (CRS), but costs are lower the larger the industry in which the firm is located (e.g., well-developed infrastructure and a large supply of skilled workers consequent on a large industrial sector). Internal economies are inconsistent with a perfectly competitive equilibrium, and external economies are, well, externalities. Since the theory of compara-

tive advantage assumes perfect competition and no externalities, trade due to economies of scale alone *cannot* be comparative-advantage trade.

National versus international. Economies of scale may depend on the scale of operations within a nation (e.g., large plant size) or on the scale of operations globally (e.g., division of labor and free trade in intermediate goods). Either might be internal or external to the firm. An example of internal, international economies of scale is research and development (R&D) by a multinational firm that utilizes the results of the R&D in several countries.

Aggregative versus disaggregative. Increasing returns may be a property of manufacturing generally (e.g., the size of the industrial sector) or of individual manufactured goods (e.g., the number of red sedans).

These three considerations generate eight types of scale economies, each relevant in reality. Comparative-advantage trade can also be due to many causes, but they all matter *solely* in terms of how they influence differences in relative autarky prices. This imparts an attractive formal unity to that theory's predictions. Trade due to economies of scale is dramatically different: Basic implications are indeed very sensitive to the *type* of scale economy. Consider first national, aggregative, increasing returns external to the firm.

National, Aggregative Economies of Scale External to the Firm IRS can furnish a basis for trade independent of comparative advantage. Consider a simple model with two identical economies with two-good Ricardian technologies. Good A is produced with CRS, with one unit of labor required in each country to produce one unit of A. The B sector has IRS external to the firm: $B = k(L_B)L_B$, where $k(L_B) = k_0(L_B)^{\alpha - 1}$ for $\alpha > 1$. The individual firm takes k as a parameter.

Two such identical countries will have equal relative autarky prices. But there is still a basis for trade: with IRS in the B industry it is not globally efficient for both countries to produce both goods.

The no-trade case, with each country doing what it had done in autarky, is a free-trade equilibrium. But it will not be a very stable one. For suppose the home economy produces more B, and the foreign

economy less. Then k is greater than k^*, the foreign analog, so home B firms can undersell their foreign rivals, while the two countries can still produce A on equal terms. Home B producers can increase their market share at the expense of their foreign competitors, so foreign resources move into the A sector. Thus k rises further and k^* falls, increasing the home B advantage still more. This continues until a new equilibrium is reached with the home economy producing only B and/or the foreign economy producing only A. Indeed there is more than just a second equilibrium: since the two countries are identical we can find a third equilibrium by simply reversing the home and foreign roles.

The basic idea behind comparative advantage is that countries should do what they can do relatively well; this implies some particular role in the world economy. Scale economies on the other hand require countries to concentrate on a small number of tasks; who does what is secondary. Thus scale economies introduce a bias toward a multiplicity of equilibria.

The other equilibria might involve one country specialized in B and/or one specialized in A. Consider an equilibrium in which the home country specializes in B, with the (identical) foreign economy producing both goods. Since $k > k^*$, foreign wages must be lower than home wages and also lower than wages in autarky, so foreign real income must have declined. Likewise, the home wage must have risen relative to autarky. Thus trade has benefited the home economy, but has made the foreign worse off. There is also a "mirror-image" equilibrium in which the roles of the two countries reverse. Thus potential international conflict is inherent. We might call this the Graham case since it corresponds to Graham's (1923) argument for protection.

Suppose now that, instead, the dynamic adjustment ends with the foreign economy specialized in A and the home economy producing both goods. Since both countries produce A, wages must be the same internationally, in sharp contrast to the Graham case. Thus residents of both countries fare the same, and that common fate must be an improvement over autarky, since the home B sector has grown. There will again be another "mirror image" equilibrium,

but unlike the previous case this is of no consequence, because everyone fares the same regardless of country of residence. With identical economies, a wage-equalization equilibrium is associated with a large world equilibrium demand for A, so that one country alone cannot satisfy it. But with dissimilar countries, it is easy to construct examples in which either the larger or the smaller country specializes in A, and in which both countries lose relative to autarky.

The final possibility is that both countries specialize. Then the international equilibrium is efficient, unlike the other cases, where too little B is produced. The various types of equilibria are not mutually exclusive. That is, if tastes, technology, and size imply multiple equilibria, the equilibria could be of different types. It is the possibility of wage equalization and (especially) Graham equilibria that produces the real value added that can come from consideration of IRS. These equilibria can produce positive and normative implications in sharp contrast to those of comparative advantage and can therefore be used in support of quite different policy recommendations. They are of direct relevance to the old debates in developing countries over the wisdom of participating in the international trading system.

But this analysis of national external economies of scale is less than fully satisfying and has accordingly had to play a role very much subservient to that of comparative advantage. The indeterminacy of results due to the likelihood of multiple equilibria renders the theory cumbersome to use. Also, this investigation of scale economies was motivated in large part by a desire to address more directly a world in which the lion's share of trade consists of the exchange of similar commodities between similar economies. But the influence of scale economies, enhancing the possibility of specialization, and perhaps causing initially similar economies to become very dissimilar, is to move the discussion in just the opposite direction. Additional methods of modeling scale economies are needed.

Disaggregative Economies of Scale Consider now the same model as above, *except* that the B sector now consists of n distinct varieties, B_i, each with the technology described above. Note the following concerning international trade in such a model.

1. Wage equalization equilibria will again feature both countries producing some A, but no variety of B in common, so that both are equally well off with free trade. If one country specializes in A all trade will consist of the *inter*industry exchange of A for B, but if both countries produce some varieties of B there will also be an *intra*industry exchange of B varieties. We would expect the latter to be relatively more important the more similar the two countries are: if the two are exactly alike, there will be an equilibrium with only intraindustry trade.

2. Graham equilibria can still emerge whenever the two countries produce some variety of B in common but in different amounts. The country with the larger production must have the higher wage and therefore cannot be producing any A. However this now seems like a much less likely outcome than before, since a smaller B sector than in autarky need not condemn a country to a lower wage: it can just produce a smaller number of varieties while fully supplying the world demand for each.

International Economies of Scale External economies have often been identified with an increased division of labor made possible by a larger market: Adam Smith's pin factory and the Swiss watch industry are the prominent hoary examples. Less common are examples having to do with a larger volume of public information generated by a larger industry. In principle none of these requires an industry to be physically located in one place. A dispersed industry can realize a great division of labor if intermediate components can be shipped from place to place; public information can be dispersed within the industry if communication is efficient. What matters, under these conditions, is the global size of the industry, not its geographical concentration.

This suggests that returns to scale may depend on the size of the *world* industry, not the national industry. This is what is meant by *international* returns to scale. Suppose that resources are used to produce A and m. A production is characterized, as in the above

one-factor models, by CRS; m is an index of the scale of operations of the national B industry, subject to IRS. With *national* returns to scale, national B production B is related to m by

$B = km$ where $k = k(m), k' > 0.$

With *international* returns to scale, on the other hand, we have instead

$B + B^* = k(m + m^*)$ where
$k = k(m + m^*), \ k' > 0.$

Here an asterisk refers to the foreign country.

At first glance it might seem that we have complicated matters enormously. National production-possibility frontiers between final goods are not even defined, because productivity in each country's B industry depends on the size of the other country's B industry. But the situation becomes almost transparent as soon as we focus on patterns of resource allocation rather than on goods.

To see this, consider the world production-possibility frontier between A and B. A point on it can be found by maximizing world B production for a given feasible volume of world A production, that is, by choosing m, m^* to

maximize: $B + B^* = k(m + m^*)[m + m^*]$
subject to:
$T(m) + T^*(m^*) = some \ specified \ value.$

T and T^* denote the home and foreign production-possibility frontiers between A and m. Clearly, $B + B^*$ will be maximized by maximizing $m + m^*$: This problem has exactly the same solution as that of choosing m, m^* to

maximize: $m + m^*$
subject to:
$T(m) + T^*(m^*) = some \ specified \ value$

and the solutions to problems of the latter sort are just the comparative advantage predictions. Efficient patterns of world activity in A and B correspond to efficient patterns in A and m, ignoring the scale economies.

Productive efficiency is as with CRS, and firms behave competitively because the economies are external to them. The result is that the complex tendencies associated with Graham equilibria when scale economies are national disappear when they become international.

The second major implication of international IRS is that they imply a theory of the intraindustry exchange of intermediate goods between relatively similar economies. The essential idea behind international IRS is that a dispersed industry can realize the benefits of a large division of labor if intermediate goods can be shipped within the industry. Thus the more nearly equal in size m and m^* are, the greater the volume of intraindustry trade in B components.

All trade will be interindustry if the disparity between countries is great enough for the A exporter to specialize completely in A. Small international differences reduce the incentive for interindustry trade but cause the integrated B industry to be divided relatively evenly between countries, thereby inducing intraindustry trade. In the limiting case where the countries are identical, they will both be self sufficient in A. But they can gain from trade by establishing a single, rationalized B industry; all trade will be intraindustry.

Applications of Scale Economies Economies of scale can be a basis for trade, just like comparative advantage. But they can give a very different picture of the consequences of such trade. This has clearly emerged from recent work in the area.

The notion of international economies of scale, developed by trade theory, has been used to rejuvenate the theory of economic growth, as in the work of Romer (1986). If the division of labor is limited by the extent of the market, it is reasonable to suppose that indivisibilities in the production of intermediate goods is the reason why. So there is a role for imperfect competition with regard to the latter.

Imperfect competition and scale economies have been investigated in the context of both consumer goods, by Krugman (1979), Lancaster (1980), and Helpman (1981), and producer goods, by Ethier (1982a). As this literature does involve imperfect competition, it is beyond the scope of the present entry. More recently, scale economies have been central to the literature on trade with heterogeneous firms, as in the work of Melitz (2003).

Though scale economies were in trade theory from the beginning, their role was basically tangential until the late 1970s. Now it is central.

See also comparative advantage; intrafirm trade; intraindustry trade; monopolistic competition; New Trade Theory; Ricardian model

FURTHER READING

Ethier, Wilfred J. 1979. "Internationally Decreasing Costs and World Trade." *Journal of International Economics* 9 (1): 1 24. Discusses the implications of international economies of scale.

———. 1982a. "National and International Returns to Scale in the Modern Theory of International Trade." *American Economic Review* 72 (3): 389 405. Considers scale economies and trade in producer goods.

———. 1982b. "Decreasing Costs in International Trade and Frank Graham's Argument for Protection." *Econometrica* 50 (5): 1243 68. A discussion of how trade may or may not be harmful in the presence of scale economies.

———. 1998. "Regionalism in a Multilateral World." *Journal of Political Economy* 6 (106): 1214 45. An application of external, international economies of scale to regional integration.

Graham, F. 1923. "Some Aspects of Protection Further Considered." *Quarterly Journal of Economics* 37: 199 227. An early and forceful argument that economies of scale might cause trade to harm a country.

Helpman, Elhanan. 1981. "International Trade in the Presence of Product Differentiation, Economies of Scale, and Imperfect Competition: A Chamberlin Heckscher Ohlin Approach." *Journal of International Economics* 11: 305 40. A discussion of scale economies in the production of consumer goods.

———. 1984. "Increasing Returns, Imperfect Markets, and Trade Theory." In *Handbook of International Economics*, vol. 1, edited by Ronald W. Jones and Peter B. Kenen. Amsterdam: North Holland, 325 65. A useful survey of the early "new" trade literature emphasizing economies of scale.

Krugman, Paul. 1979. "Increasing Returns, Monopolistic Competition, and International Trade." *Journal of International Economics* 9 (4): 469 79. An influential discussion of scale economies in the production of consumer goods.

Lancaster, Kelvin. 1980. "Intraindustry Trade under Perfect Monopolistic Competition." *Journal of International*

Economics 10: 151 76. An early discussion of scale economies in the production of consumer goods.

Markusen, J., and J. Melvin. 1984. "The Gains from Trade Theorem with Increasing Returns to Scale." In *Monop olistic Competition and International Trade*, edited by H. Kierzkowski. Oxford: Oxford University Press, 10 33. An important treatment of the traditional gains from trade argument in the context of scale economies.

Matthews, R.C.O. 1949 50. "Reciprocal Demand and Increasing Returns." *Review of Economic Studies* 37: 149 58. A good example of the earlier literature on scale economies.

Melitz, Marc. 2003. "The Impact of Trade on Intraindustry Reallocations and Aggregate Industry Productivity." *Econometrica* 71: 1695 1725. Considers heterogeneous products and international trade.

Romer, Paul. 1986. "Increasing Returns and Long Run Growth." *Journal of Political Economy* 94: 1002 37. A major application of "international" economies of scale to economic growth.

WILFRED J. ETHIER

■ effective exchange rate

The effective exchange rate is a summary measure of the rate at which a country's currency exchanges for a basket of other currencies, in either nominal or real terms. Effective exchange rates become relevant when a country conducts trade and investment transactions with a number of other countries. These rates can vary along several dimensions, including country coverage, weighting, and whether or not the effect of inflation is taken into account (i.e., the distinction between nominal and real). The final selection of the appropriate definition and calculation of the effective exchange rate depends on a rather complicated interplay of the theoretical model of interest and data availability and reliability.

The first issue to confront in calculating an effective exchange rate is how to attribute relative weights to each of the partner currencies. In many circumstances, the exchange rate plays the role of a relative price of traded goods; hence, the relevant weights involve trade weights. Asset trade has become more prominent in recent decades, suggesting that the appropriate weights may be related to the assets owed by, or liabilities owed to, other countries.

Trade-Weighted Effective Exchange Rates By far the most common means of calculating an effective exchange rate is to weight the currencies by trade weights. To fix concepts, consider a geometrically weighted average of bilateral exchange rates.

$$s_t^{effective} \equiv \sum_{j=1}^{n} w_j s_t^j \tag{1}$$

$$q_t^{effective} \equiv \sum_{j=1}^{n} w_j q_t^j \tag{2}$$

where $s^j (q^j)$ denotes the log nominal (real) exchange rate relative to country j. The weights w_j are usually based on bilateral trade volumes (the sum of exports and imports, expressed as a proportion of total exports and imports).

Trade weighting can take on a more complicated form to allow for competition in third markets, however. This goal is usually accomplished by adopting the Armington (1969) assumption that goods are differentiated by location of production. The third market weight is equal to the weighted average over all third-country markets of country j's import share divided by a weighted average of the combined import share of all of country i's competitors, with the weights being the shares of country i's exports to the various markets. This simple expression is based on the assumption that all the differentiated goods share the same constant elasticity of substitution, which may not necessarily be appropriate in all instances (Spilimbergo and Vamvakidis 2003). For instance, goods originating from less-developed countries may not be equally substitutable with goods originating from industrial countries. Moreover, these weights can change over time continuously or discretely and infrequently, with the choice depending in large part on the trade-off between convenience and accuracy.

Nominal versus Real Often the economist will encounter a model wherein the real, or inflation-adjusted, exchange rate plays a central role. There are a number of real exchange rates, or "relative prices," that appear in the literature, however, so there is

ample scope for confusion. A decomposition of the most standard definition is useful. In this definition, the real exchange rate is given by:

$$q_t \equiv s_t - p_t + p_t^* \qquad (3)$$

where s is the log exchange rate defined in units of home currency per unit of foreign currency.

Most models of the real exchange rate can be categorized according to which specific relative price serves as the object of focus. If the relative price of nontradables is key, then using a broad price index encompassing tradables and nontradables is implied. One example of the use of a broad index is in productivity-based explanations of the real exchange rate such as the Balassa-Samuelson model. If, on the other hand, external balance (i.e., current account balance) is of paramount concern, some narrower index of traded goods may be the appropriate deflator. This variable is also what macroeconomic policymakers often allude to as price competitiveness a weaker domestic currency (in real terms) means that it is easier to sell domestic goods abroad. A related concept is cost competitiveness. Assuming a cost-markup model of pricing (i.e., prices equal some markup over cost), one can calculate a measure of the real effective exchange rate where unit labor costs are used instead of prices (Golub 1994). This real exchange rate is best thought of as a measure of the relative *production cost* rather than price of goods.

In practice, one has a choice of only a few price deflators. At the monthly frequency, they include the consumer price index (CPI), the producer price index (PPI) or wholesale price index (WPI), or the export price index. At lower frequencies, such as quarterly data, the set of deflators increases somewhat, to include the gross domestic product (GDP) deflator and price indexes for the components of GDP, such as the personal consumption expenditure deflator. Typically, the CPI weights nontraded goods such as consumer services fairly heavily. Similarly, the GDP deflator and the CPI will weight expenditures on nontradables in proportion to their importance in the aggregate economy. In contrast, the PPI and WPI exclude many retail sales services that are likely to be nontraded.

The unit labor cost deflated index is in a sense the most relevant for many issues related to trade, as unit labor costs are a measure of cost competitiveness. Unfortunately, there are many difficulties with using such indexes. First, unit labor costs (ULCs) are not always available on a timely or consistent basis, and are subject to substantial revisions. Second, their greater covariation with the business cycle impedes discerning trends in the ULC deflated series. Third, measured ULCs usually pertain only to manufacturing sectors, and given the increasing tradability of services, measured ULCs may provide misleading inferences. Fourth and perhaps most important, ULCs are typically available on a consistent basis for developed economies, so that ULC deflated effective exchange rates can be calculated only against a reference group of countries that may not, in the end, be the relevant group.

Asset and Liability Weights In the preceding discussion, it has been taken as a given that the appropriate weights are those associated with trade flows. Yet there is no reason why trade weighting should be appropriate for all questions. The economists Cedric Tille, Hélène Rey, and Pierre-Olivier Gourinchas have pointed out that exchange rate changes have had substantial effects on the net international investment position of the United States. Tille (2003) noted that because U.S. assets are predominantly denominated in foreign currencies, while U.S. liabilities are mostly denominated in dollars, dollar depreciation induces a large upward effect on the dollar valuation of U.S. foreign assets. Hence, over the short to medium term, the net international investment position is heavily influenced by dollar movements.

See also Balassa-Samuelson effect; band, basket, and crawl (BBC); equilibrium exchange rate; exchange rate forecasting; exchange rate regimes; exchange rate volatility; nontraded goods; purchasing power parity; real exchange rate

FURTHER READING

Armington, P. S. 1969. "A Theory of Demand for Products Distinguished by Place of Production." *IMF Staff Papers* 16: 159 78. Provides the basis for the modeling of trade flows.

Bayoumi, Tamim, Jaewoo Lee, and Sarma Jayantha. 2006. "New Rates from New Weights." *IMF Staff Papers* 53 (2): 272 305. Describes the most recent incarnation of the IMF indexes.

Black, Stanley W. 1976. "Multilateral and Bilateral Mea sures of Effective Exchange Rates in a World Model of Traded Goods." *Journal of Political Economy* 84 (3): 615 22. One of the earliest descriptions of how to con struct effective trade weighted exchange rate indexes.

Chinn, Menzie. 2006. "A Primer on Real Effective Ex change Rates: Determinants, Overvaluation, Trade Flows, and Competitive Devaluations." *Open Economies Review* 17 (1) (January): 115 43.

Golub, Stephen. 1994. "Comparative Advantage, Exchange Rates, and Sectoral Trade Balances of Major Industrial Countries." *IMF Staff Papers* 41 (2) (June): 286 313. Links the idea of cost competitiveness to effective ex change rates.

Gourinchas, Pierre Olivier, and Hélène Rey. 2005. "Inter national Financial Adjustment." NBER Working Paper No. 11155 (February). Cambridge, MA: National Bu reau of Economic Research. An important paper that implements the construction of asset weighted effective exchange rate indexes.

Lane, Philip, and Gian Maria Milesi Ferretti. 2005. "Fi nancial Globalization and Exchange Rates." IMF Working Paper No. 05/03 (January). Washington, DC: International Monetary Fund. This paper notes that in a world characterized by increasing financial globalization (namely increasingly large cross border holdings of as sets and debt) exchange rate movements will have ever more pronounced effects on *net* asset positions in many countries, in addition to the United States.

Spilimbergo, Antonio, and Athanasios Vamvakidis. 2003. "Real Effective Exchange Rate and the Constant Elas ticity of Substitution Assumption." *Journal of Interna tional Economics* 60 (2): 337 54. This paper drops the assumption that goods from developed and less devel oped countries are not equally substitutable.

Tille, C. 2003. "The Impact of Exchange Rate Move ments on U.S. Foreign Debt." *Current Issues in Eco nomics and Finance* (Federal Reserve Bank of New York) 9 (1): 1 8.

MENZIE D. CHINN

■ effective protection

Effective protection measures the change in value added per unit of output induced by tariffs and other policy measures on competing imports and inter mediate inputs. The concept has been widely used by trade economists and policy analysts interested in investigating trade policies' protective effects on do mestic production activities. They recognize that trade policies affect both the price of a good produced domestically and the price paid for any inputs in volved in producing that (final) good. The gross price of the final good is raised by a tariff on competing imports, giving rise to apparently positive protection in nominal terms and an incentive to increase pro duction. The net price or value added in the activity may still be reduced even where there is positive nominal protection, however, if the price-raising effects of tariffs on intermediate inputs more than offset the effects of the final output tariff.

Economists had long recognized the danger of focusing on output tariffs to analyze the protective effects of a tariff structure, but it was Max M. Corden (1971) who systematically organized his and others' seminal contributions during the 1960s into an in tegrated theory of effective protection.

Modeling of the Concept Within a partial equilibrium modeling framework with fixed tech nologies (no scope to alter the mix of factor and material inputs) and tradable inputs only, effective protection can be measured using information on output (t_j) and input (t_i) tariff rates and technological or input share coefficients under free trade conditions (a_{ij}). The resulting measure of effective protection (e_j) summarizes the impact of the tariff structure on value added in each activity and indicates the corre sponding production and resource allocation effects. For the single input and single output case:

$$e_j = \frac{t_j - a_{ij} t_i}{1 - a_{ij}} \qquad (1)$$

From equation (1) some key insights of the con cept can be derived. Effective protection is positively related to the output tariff and negatively related to the input tariff output tariffs implicitly subsidizing activity j and input tariffs taxing it. Effective pro tection can be positive or negative depending on the

structure of protection (whether output tariffs are greater or smaller than input tariffs) and on the shares of intermediate inputs. Further, the effective rate of protection will exceed the nominal tariff rate if there is an escalating tariff structure, with output tariffs higher than input tariffs, the effects of tariff escalation being magnified the higher the share of intermediate inputs (or correspondingly the lower the degree of domestic value added). Indeed, with sufficiently steep tariff escalation and highly (imported) inter-mediate-dependent activities, value added at world prices (i.e., in the absence of tariffs) can be negative.

Measurement and Application The effective protection concept seeks to provide a framework for thinking systematically about the structure of tariffs and other trade policy measures, and the resulting structure of protection and incentives. It offers a coherent explanation of why tariff escalation is a widely observed feature of trade regimes. It offers an explanation also of why the degree of actual and perceived protection may vary substantially across different types of activities and sectors. Since a given product (e.g., steel) can be a final product of one set of producers (steel producers) and an input good for another set (e.g., car manufacturers), it is not possible to ensure an escalating tariff structure for all pro-ducers. Therefore, the effective rate of protection is an analytical tool that offers an explanation for both idiosyncratic features of trade policy regimes and systematic biases: the possibility of high rates of ef-fective protection for many import-substituting ac-tivities but, at the same time, low negative rates for specific activities. Similarly, evidence has been found of systematic biases in favor of final goods production and against intermediate goods, as well as in favor of import-substitution activities and against export ac-tivities, both of which can be understood and quantified using the effective protection framework.

Following Corden's developmental work, a gen-eration of applied trade economists sought to mea-sure effective protection (mainly within the partial equilibrium modeling framework described earlier) and to use these measures for trade policy appraisal purposes, in particular but not exclusively in a de-veloping country context. Its tractability was attrac-tive for applied work. It can be applied at the firm or industry level, and it is relatively straightforward to allow for the complexities of multi-input and output activities, factor in nontraded inputs, distinguish sales of final goods to protected domestic markets from those to unprotected export markets, and ac-count for a range of taxing and subsidizing effects induced by policy instruments other than tariffs. Compared with applied and economywide general equilibrium models, partial equilibrium estimates have relatively limited data requirements. In addition to tariff data (and possibly other policy information), the main data requirement is information on the technology or input share coefficients available from input-output tables or firm-level surveys.

Following the major multicountry studies by Anne O. Krueger (1978) and Bela Balassa (1982), there was an explosion of developing country studies that measured and analyzed effective protection within a partial equilibrium framework (reviewed by Greenaway and Milner 1993). Although there has been a significant liberalization of trade policies in developing countries since the late 1980s, effective rates of protection are still measured and analyzed, including in the World Trade Organization's Trade Policy Reviews. There has also been a renewal of interest in the role of transportation and other non-trade policy sources of international transaction costs. These costs also have implicit taxing and sub-sidizing effects that can be represented and analyzed within an effective protection framework.

Theoretical Criticisms From the outset trade theorists, such as Ronald Jones (1971), who were interested in whether the interpretation of partial-equilibrium measures survived in general equilib-rium, closely scrutinized the concept of effective protection. A key assumption of the partial equilib-rium approach is fixed technological coefficients, and therefore no substitutability among material inputs and between material and factor inputs. Defining and interpreting changes in value added becomes more problematic if substitution is possible. This, and the wider limitations of using differences be-tween products or sectors in the scale of measured effective protection alone to comment on general

equilibrium resource pulls, led some economists to conclude that the concept was "fatally flawed." Others have been more willing to distinguish between the theoretically possible and the probable problems in practice. Some applied general equilibrium (AGE) modeling has compared the resource allocation effects of tariffs under alternative substitution assumptions, and the rank correlations between effective protection measures and tariff-induced production changes are not necessarily low. Effective protection measures do not necessarily, therefore, provide poor guidance on the allocative effects of tariffs. Further, in a review of the arguments, Greenaway and Milner (2003) point out that many applied economists have been careful to avoid interpreting differences between sectors or changes over time in measured effective protection as providing detailed and precise guidance on resource allocation effects. Rather, they have used the overall pattern of estimates to comment on the likely systematic biases and distortions of a trade policy regime and to offer guidance on the general direction of policy reform.

The development and refinement of economy-wide AGE models, and the associated developments in software that have eased their construction, calibration, and solution, have extended their use in trade policy evaluation. Where they are available, AGE models clearly give direct information on the resource allocation effects of trade policy changes. But the data demands of such models are high, even at the relatively high levels of product aggregation typically involved. The effective protection concept, and estimates of effective protection interpreted judiciously, will continue to be useful for analyzing trade policy regimes and their potential effects on incentives, resource allocation, and incomes. This will help to shed light on patterns of protection in the world economy.

See also applied general equilibrium modeling; partial equilibrium modeling; tariff escalation; tariffs

FURTHER READING

Balassa, Bela, ed. 1982. *Development Strategies in Semi Industrialized Economies.* Baltimore: Johns Hopkins University Press. A helpful summary of the empirical issues associated with measuring effective protection, with evidence for several industrializing countries.

Corden, Max M., ed. 1971. *The Theory of Protection.* Oxford: Oxford University Press. The first systematic exposition of the theory of effective protection.

Greenaway, David, and Chris R. Milner, eds. 1993. *Trade and Industrial Policy in Developing Countries.* London: Macmillan. Reviews the tools used to analyze trade policy and policy reform in developing countries.

———, eds. 2003. "Effective Protection, Policy Appraisal, and Trade Policy Reform." *The World Economy* 26 (4): 441 56. Reviews the empirical and theoretical literatures and offers a defense of the use of effective protection in applied work.

Jones, Ronald W., ed. 1971. "Effective Protection and Substitution." *Journal of International Economics* 1 (1): 59 82. An early example of the theoretical criticism of partial equilibrium measures of protection.

Krueger, Anne O., ed. 1978. *Liberalization Attempts and Consequences.* New York: National Bureau of Economic Research. A commonly cited, cross country study of the costs of protection in developing countries.

CHRIS MILNER

■ electronic commerce

Electronic commerce, or *e-commerce*, has no settled definition. At its broadest, electronic commerce involves conducting business using most modern communication instruments: telephone, fax, television, electronic payment and money transfer systems, electronic data interchange, and the Internet. On September 25, 1998, the World Trade Organization (WTO) General Council adopted a broad view of electronic commerce in its work program on the subject: "The production, distribution, marketing, sale or delivery of goods and services by electronic means." In more recent times, the term *e-commerce* has become strongly associated with commercial activities on the Internet. For instance, the Organisation for Economic Co-operation and Development's discussions of e-commerce concentrate almost exclusively on Internet-based transactions.

341

In this article, electronic commerce is conceived of as conducting or facilitating business via electronic communications networks and computer systems. This includes buying and selling online, electronic funds transfer, business communications (including by telephone, facsimile, and internal data networks), and using computers to access business information resources. The WTO has recognized that commercial transactions can be broken into three stages: (1) advertising and searching, (2) ordering and payment, and (3) delivery. Common conceptions of electronic commerce involve business-to-consumer or business-to-business interaction at one or more of these three stages. Generally speaking, however, electronic commerce also encompasses activities that do not fit neatly into any one of these categories, such as electronic logistics tracking and business process outsourcing.

The Internet is of such critical importance to the world economy today that it necessarily dominates any discussion of electronic commerce. It offers greater possibilities for commercial interaction than do telephones or faxes, and has dramatically changed the way that information is exchanged and business conducted. The definition of electronic commerce adopted here deliberately extends beyond the Internet, however, not only to capture its predecessors, but also in anticipation of future technologies. If we have learned anything from the rapid development of computer and telecommunications technology over the last few decades, it is that what seems like an established technology today can quickly be replaced by another technology. The safest prediction about the future direction of technology is that it will be unexpected. Thus, as ubiquitous and permanent a feature of modern life as the Internet seems now, it may be replaced, and it is therefore appropriate to use a definition that is technologically neutral and forward looking.

The Economic Significance of Electronic Commerce Three factors contribute to the significance of electronic commerce to the world economy today: the rapid growth of the Internet, its ability to facilitate cross-border trade, and its ability to reduce transaction costs.

Although it became publicly accessible only with the inception of the World Wide Web in 1990, the Internet now has more than 1 billion users. In the early years of the World Wide Web, usage almost doubled from year to year, and it continues to grow. The benefit of conducting business online increases exponentially with the number of connections.

The Internet facilitates production and distribution across borders so that, for example, consumers in one country can respond to an advertisement published in another country for a product that was developed using a design team collaborating (using the Internet) in five other countries. Electronic commerce increases the range of services that can be traded internationally (e.g., medical, legal, educational, and gambling services) and can assist in opening markets that were previously closed. The dramatic increases in both online retail sales and advertising revenues are the most visible evidence of a much broader growth in electronic commerce in the global economy.

Electronic commerce has the potential to generate benefits beyond those of trade liberalization on its own. Benefits for suppliers and vendors include reduced transaction costs, reduced barriers to market entry, more rapid product innovations, and economies of scale. One source of reduced transaction costs is the possibility to dispense with traditional intermediaries (for example, in relation to the travel industry). In many cases, electronic commerce dispenses with the need for physical presence at the point of sale or for the provision of services, which can drastically reduce expenses relating to premises or personnel. Benefits for consumers include increased market transparency and reduced search costs, even if they make the final purchase in person. These benefits are particularly noticeable for consumers in smaller markets, who may not have enjoyed the same level of price and quality competition as consumers in larger markets. Challenges to increased consumer use of electronic commerce include concerns about information privacy and fraud, which have not been completely resolved. Governments also benefit from electronic commerce because it reduces the cost of

providing services to their citizens, while offering greater transparency and accountability.

These three factors have certainly altered the conditions in many preexisting markets for goods and services, but two markets are particularly useful in demonstrating these factors. First, consumer-to-consumer commercial transactions, facilitated by accessible online financial services, community, and auction sites, now have the potential to operate on a global, cross-border level. It is difficult to draw a bright line between consumer-to-consumer and business-to-consumer transactions, as the removal of many barriers to entry have allowed some consumers to play a role that was once available only to established global players. This reveals one reason for the exponential benefit of growth in the Internet's user base: unlike previous communications technologies, which facilitated one-to-one or one-to-many marketing and sales, Internet-based e-commerce allows a many-to-many business paradigm to become practicable.

Second, the market for digital products and services (those that can be supplied to the purchaser electronically) is arguably the most substantially affected by the growth of electronic commerce. Once they are provided in a digital format, goods such as music, movies, and software have negligible production costs for additional units and equally negligible delivery costs. Over the Internet, their delivery time is limited primarily by bandwidth and connection speed rather than distance or the speed of physical transport. Services that can now be provided digitally across borders are increasingly being outsourced to take advantage of cheaper labor markets. This phenomenon, called business process outsourcing, began in its modern incarnation with the outsourcing of software development to India. India remains the market leader in business process outsourcing exports today and has moved into many other areas such as sales and customer service call centers as well as intrabusiness services for example, information technology support and human resources. Other developing countries, such as Brazil and China, have also experienced strong growth in this area. As the uptake and speed of Internet con-

nections increase, the sale of digital products and services is likely to have a substantial impact on the world economy, not only in its own right, but also due to its disruptive effects on related "offline" industries.

Conversely, many factors militate against electronic commerce completely replacing traditional nonelectronic business models. Thus, despite its growth, the uptake of electronic commerce was slower than many expected. As mentioned earlier, the security of online payment and information transfer remains a primary concern. The market for digital products is especially affected by the lack of cost-effective micropayment methods (i.e., for low-value transactions). Additionally, due in part to the threat of fraud, or possibly more for sociological than economic reasons, many consumers and businesses prefer physical presence over electronic transactions, especially to conduct a physical inspection of goods before purchase. This has led to the perception that online retailers "freeload" off retailers with physical stock and showrooms, as customers can investigate the product in person in a retail store and then purchase it from an online store at a reduced price. Some manufacturers and suppliers refuse to sell to online retailers in order to protect their existing showroom-based retail outlets. Furthermore, although the growth of the Internet has been substantial, many use the Internet for recreational and communication purposes rather than commercial ones, and a majority of the world's population is still without regular Internet access.

Nevertheless, as electronic commerce matures and develops, it will undoubtedly play an increasingly important role in the world economy. Many of the impediments that it is presently facing are technological in nature, and given the speed at which new developments occur in information technology, it is only a matter of time before many of them are overcome.

The Role of the WTO Although electronic commerce involves new technology, at its heart it is simply another means of conducting international commercial transactions. Most countries have previously decided that such transactions should be dealt

with in the WTO. The advantage of locating electronic commerce within the WTO framework is that it provides a system of transparent, predictable, and enforceable rules. These rules are based on principles of nondiscrimination (most-favored-nation treatment [MFN] and national treatment) and transparency, which are as relevant to electronic commerce as they are to other forms of international trade. The WTO also has an established record of trade liberalization and takes an economic and commercial focus. This means that it is not particularly concerned with technology matters; rather it is intent on creating a system of trade agreements that apply to transactions regardless of the form of technology used to produce or deliver the product. Many WTO members share this belief in the principle of technological neutrality. It is important to scrutinize deviations from neutrality: for example, those resulting from the goods and services distinction created by the General Agreement on Tariffs and Trade (GATT) and the General Agreement on Trade in Services (GATS). Technological neutrality should not be used as an excuse for protectionism, however. The primary purpose of the WTO remains trade liberalization. Thus, for example, practices that liberalize trade in a good only where it has been created using a certain technology (i.e., technologically discriminatory liberalism) should be preferred to practices that create or maintain barriers to trade in that good however created (i.e., technologically neutral protectionism).

On May 20, 1998, during its second Ministerial Conference in Geneva, the WTO adopted a Declaration on Global Electronic Commerce, recognizing the growth of electronic commerce and its potential to increase international trade. The declaration directed the General Council to establish a comprehensive work program to examine all trade-related issues of global electronic commerce and to produce a report on the progress of the work program and any recommendations for action at the third Ministerial Conference of the WTO. The Council for Trade in Goods, the Council for the Agreement on Trade Related Aspects of Intellectual Property Rights (TRIPS), the Committee on Trade and De-

velopment, and the Council for Trade in Services each produced a report for the General Council in July 1999. The general view of members emerging from these and subsequent reports is that the electronic delivery of services falls within the scope of GATS, and that all the provisions of GATS apply to trade in services through electronic means. Members have not reached a consensus on issues such as the treatment of certain digitizable products, however, and work continues on resolving these issues. The General Council has held a number of dedicated discussions on "cross-cutting" e-commerce issues, in recognition of the fact that many e-commerce problems cut across a number of WTO agreements such as the GATS, GATT, and TRIPS. These discussions have covered topics such as the classification of digital products and the fiscal implications of e-commerce. The Councils for Trade in Goods, Trade in Services, and TRIPS as well as the Committee on Trade and Development have also held discussions on e-commerce issues relating to their respective mandates.

The general view of WTO members that GATS applies to electronic delivery of services and trade in services through electronic means is consistent with the prevailing academic view that electronic commerce, for the most part, falls under the purview of GATS. Many WTO members have expressed the view that all digital products should be classified as services so as to fall under GATS. GATS extends the scope of international trade obligations to cover services, which represent a significant and growing proportion of global trade. It imposes general principles of MFN treatment and transparency to services (subject to some exceptions), and it also provides for individual commitments by members to liberalization of trade in services, including commitments to provide market access and national treatment in specified service sectors.

Once WTO members agree on how to handle electronic commerce, WTO agreements and dispute settlement have the potential to profoundly aid or curtail the growth and impact of e-commerce, especially as regards cross-border trade. Other international bodies will also have an effect on the future

development of e-commerce, for instance the UN Working Group on Internet Governance. The proven effectiveness of the WTO in securing a degree of multilateral compliance with its international trade standards, however, makes it a key player in the regulation of electronic commerce.

Trends in Electronic Commerce A number of trends are evident in relation to electronic commerce that will affect the world economy. First, the value and number of digital products (e.g., music, books, and videos delivered electronically rather than physically) is likely to continue to increase. Second, communications technologies are gradually converging. If this trend continues, previously separate technologies such as the telephone, television, and the Internet may eventually be considered part of a single communications network. This would result in traditional media companies (such as television broadcasters) being compelled to shift to a more interactive, Internet-based model, and online vendors would benefit from greater reach to mobile and entertainment devices. Third, the Internet is continuing to affect consumer behavior. For example, consumers are spending more time and money online, and the value of their online purchases is increasing as they gain experience with online shopping and move to higher speed Internet. Fourth, transactions using notes and coins will continue to decline in favor of transactions using digital money. (Notes and coins make up only a small fraction of the total value of economic transactions, but they still comprise a significant proportion of the number of transactions.) Fifth, the growth in the business process outsourcing market has the potential to alter materially the structures and costs associated with many conventional businesses. Outsourcing specialized tasks to low-cost providers is becoming increasingly feasible, if not necessary, to compete on a global scale. This service, provided mainly by developing countries with cheaper labor, also has the potential to aid in integrating such countries into the global trading system.

The Role of Electronic Commerce in the Modern World Economy Electronic commerce will continue to play a significant role in the modern world economy in the coming years. Its potential will be limited, however, in the absence of a comprehensive understanding within the WTO or some other multilateral framework regarding how to treat international trade conducted using electronic commerce. So, too, may technological and regulatory limitations regarding issues such as privacy, financial security, and fraud hinder the development of electronic commerce beyond the level it has reached today. Experience with previous technologies and negotiations suggests that these hurdles will be overcome, however, particularly to the benefit of developing countries.

See also digital divide; General Agreement on Tariffs and Trade (GATT); General Agreement on Trade in Services (GATS); information and communication technology; trade in services; World Trade Organization

FURTHER READING

Borenstein, Severin, and Garth Saloner. 2001. "Economics and Electronic Commerce." *Journal of Economic Per spectives* 15 (1): 3 12. An effective introduction to many of the economic issues surrounding electronic commerce and to a symposium on economics and electronic commerce.

Mann, Catherine, Sue Eckert, and Sarah Knight. 2000. *Global Electronic Commerce: A Policy Primer.* Washing ton, DC: Institute for International Economics. Aimed at assisting policymakers, particularly those in develop ing countries, with formulating an appropriate response to e commerce and to use it to facilitate development.

Organisation for Economic Co operation and Develop ment, Working Party on the Information Economy. *Online Payment Systems for E Commerce.* OECD Doc DSTI/ICCP/IE(2004)18/FINAL (18 April 2006). Analyzes the recent development of online payment systems for e commerce, covering different payment mechanisms, the extent to which these different systems are used, and the implications of industry characteristics and network effects. It discusses drivers and impedi ments to the uptake of payment systems and identifies some policy issues for further examination.

UN Information and Communication Task Force. 2005. *WTO, E Commerce, and Information Technologies: From the Uruguay Round through the Doha Development*

Agenda (Series 7). Prepared by Sacha Wunsch Vincent, this paper provides a comprehensive analysis of the WTO's engagement with e commerce as well as some useful research into key areas of electronic commerce.

World Trade Organization, Secretariat. 1998. *Electronic Commerce and the Role of the WTO*. Lays out the relevant policy issues for electronic commerce and the WTO, including the legal and regulatory framework for Internet transactions, security and privacy questions, taxation, access to the Internet, market access for suppliers over the Internet, trade facilitation, public procurement, intel lectual property questions, and regulation of content.

Wunsch Vincent, Sacha. 2006. *The WTO, the Internet, and Trade in Digital Products: EC U.S. Perspectives*. Oxford: Hart. A neutral and thorough reference work examining European Community and U.S. perspectives on digital products (that is, digitally delivered movies, music, and software), and the response of international trade rules, in particular those of the WTO.

ANDREW D. MITCHELL

■ **endogenous growth theory**

See growth in open economies, Schumpeterian models

■ **equilibrium exchange rate**

Ever since the advent of floating exchange rates for the principle currencies in 1973, exchange rate misalignment has been an important issue for academics, practitioners, and policymakers. In flexible exchange rate regimes, exchange rates can spend long periods away from their equilibrium, or fair value, levels (that is, they can take long swings). Consequently, in order to understand whether exchange rates are misaligned or not, analysts need an underlying measure of equilibrium. How then may the equilibrium exchange rate be defined?

Purchasing Power Parity (PPP) Reconsidered
The equilibrium measure that many economists and policymakers first turn to is the concept of purchasing power parity (PPP). The PPP exchange rate is the ratio of some overall measure of domestic prices

relative to a comparable measure of foreign prices. Although there is controversy in the PPP literature regarding the correct measure of overall prices to use (e.g., consumer, wholesale, or some other measure) there is no longer a question about the "time dimension" of PPP. By this we mean how quickly exchange rates are expected to gravitate, or revert, to the PPP defined rate (MacDonald 2007).

Proponents of traditional PPP would argue that disturbances to PPP should be rapidly offset. But how rapidly should this be to be consistent with PPP? The classic disturbance that affects the relationship between the nominal exchange rate and relative prices is a liquidity disturbance, such as a monetary expansion. The latter, in the context of an economy in which there are sticky commodity prices, will not have an immediate effect on relative prices but would have their effect 18 months to 2 years thereafter. In this setting, therefore, a monetary shock is expected to move the real exchange rate in the short run, but the real exchange rate should revert to its mean value after around two years: a proponent of traditional PPP would argue that the so-called half-life of mean reversion would be one year.

There is now a huge empirical literature that indicates that the so-called half-life has a range of between three and five years, which is inconsistent with a traditional form of PPP; Rogoff (1996) labeled this the PPP puzzle. It is a puzzle because the stylized fact of the high volatility of real and nominal exchange rates is consistent with the interaction of liquidity shocks with sticky prices, but as we have seen, the slow mean reversion of real exchange rates is not.

Various attempts have been made to explain the PPP puzzle. One explanation, which is consistent with traditional PPP, is that the existence of transaction costs imparts a nonlinear process to exchange rate behavior, and when such nonlinear behavior is accounted for, real exchange rates behave in a manner that is consistent with traditional PPP they have a half-life of around one year. Such nonlinear explanations, however, are really nothing more than black box interpretations and are equally consistent with other interpretations of the PPP puzzle (MacDonald 2007). A second key explanation of the PPP puzzle,

which is not consistent with traditional PPP, is that real factors, such as productivity differences (usually motivated in terms of the Balassa-Samuelson effect) and a country's net foreign asset position, ultimately push real exchange rates away from their equilibrium values. A third key explanation of the PPP puzzle is the pricing-to-market behavior of multinational firms. Such firms, by altering their markup (in order to protect market share) as nominal exchange rates change, impart prolonged and persistent deviations from PPP. Although this latter view appears to capture an important determinant of the systematic behavior of real exchange rates, the second explanation may be more critical to understanding equilibrium exchange rates and, by implication, whether exchange rates are misaligned or not. We now consider these alternatives, all of which are based on an explicitly "real" interpretation of real exchange rate behavior.

Fundamental Equilibrium Exchange Rate The fundamental equilibrium exchange rate (FEER), developed by the economist Williamson (1983, 1994), is the rate at which both the internal and external balance are satisfied, where the internal balance is high employment and low inflation and the external balance is characterized as the sustainable desired net flow of resources between countries when they are in internal balance. The latter is usually derived by making some assumption about a country's net savings position, which, in turn, is determined by factors such as consumption smoothing and demographic changes. The use of the latter assumption, especially, has meant that the FEER is often interpreted as a normative approach and the calculated FEER is likely to be sensitive to the choice of the sustainable capital account. (The International Monetary Fund variant of the FEER approach attempts to remove the judgmental element in defining the desired capital account; see, e.g., Isard and Faruqee 1998; and Faruqee, Isard, and Masson 1998.) The FEER is an explicitly medium-run concept, in the sense that it does not need to be consistent with stock-flow equilibrium (the medium run is usually taken to be a period of about five years in the future).

There are essentially two widely used approaches to estimating a FEER. The first involves taking an estimated multicountry, macroeconometric model, imposing internal and external balance, and solving for the real exchange rate, which is then classified as the FEER. Such an approach is not very tractable and, consequently, by far the most popular method of generating a FEER involves focusing on a current account equation, setting it equal to a sustainable capital account, and then solving for the FEER (see Wren-Lewis 1992).

In addition to the difficulty in measuring a sustainable capital account, the calculation of trade elasticities has often meant that an extra layer of judgment has to be imposed before the FEER can be calculated. This is because the estimated trade elasticities often turn out to be effectively zero (see Goldstein and Khan 1985). Driver and Wren-Lewis (1999) assess the sensitivity of FEER calculations of the U.S. dollar, Japanese yen, and German mark to different formulations and assumptions. They find that two key factors impart a considerable amount of uncertainty into FEER-type calculations. For example, changes in the assumed value of the sustainable capital account (as a proportion of gross domestic product, GDP) of 1 percent can produce changes in the value of the FEER of around 5 percent. Since such changes in the capital account could easily be due to measurement error, this suggests caution in interpreting point estimates of the FEER. Although the FEER approach has been widely used by practitioners, the issues mentioned above often mean that it is used in conjunction with other measures of equilibrium.

Behavioral Equilibrium Exchange Rate The behavioral equilibrium exchange rate (BEER) approach of Clark and MacDonald (1998) is not based on any specific exchange rate model and, in that sense, may be regarded as a very general approach to modeling equilibrium exchange rates. As in the FEER-based approach, it takes as its starting point the proposition that real factors are a key explanation for the slow mean reversion to PPP observed in the data. In contrast to the FEER-based approach, the specific modus operandi of the BEER is to produce measures

of exchange rate misalignment that are free of any normative elements and in which the exchange rate relationship is subject to rigorous statistical testing.

To illustrate their approach, Clark and MacDonald (1998) take the risk-adjusted real interest parity relationship, which has been used by a number of researchers to model equilibrium exchange rates (see, e.g., Faruqee 1995; MacDonald 1997), where it is assumed that the systematic component of the real exchange rate is a function of net foreign assets, a measure of relative productivity, and the terms of trade. The approach is estimated using a vector error correction mechanism (VECM), which incorporates both dynamic interactions and long-run effects, and plausible measures of equilibrium have been reported for a number of real effective exchange rates (plausible in the sense that coefficients in the long-run relationship are correctly signed and statistically significant and also fast adjustment to equilibrium is reported).

Comparing the actual real effective exchange rate with the computed equilibrium gives a clear and sharp measure of misalignment. For example, the approach shows that the U.S. dollar was massively overvalued in the period 1980 86. It is worth noting that this finding is common to other BEER estimates (see, e.g., Faruqee 1995; MacDonald 1997; Stein 1999). It is also possible to calibrate the BEER with some normative structure placed on the fundamentals, much as in the FEER approach. The big advantage of the BEER, however, is that the influence of the normative elements is explicit as is the nature of the estimated exchange rate model and the adjustment to equilibrium. The BEER approach has been widely and successfully used for industrial countries, developing countries, and emerging markets.

Natural Real Exchange Rate The natural real exchange rate (NATREX) approach of Stein (1994, 1999) may be considered to be grounded in elements of the FEER and the BEER. It has similarities with the FEER because it appeals to an internal-external balance approach to motivate the model, and shares with the BEER what is essentially a reduced form approach in the empirical estimation. In the NA-

TREX, the sustainable capital account term is assumed equal to social saving less planned investment, where the key determinant of social savings is the rate of time preference and the key determinant of investment is Tobin's q. Additionally, savings are assumed to be a function of net foreign assets and investment a function of the capital stock, k. The inclusion of stocks in the flow relationships enables an equilibrium to be derived that is stock-flow consistent.

Stein (1999) proposes two forms of NATREX equilibrium. In "long-run" equilibrium the following criteria have to be satisfied. First, net foreign assets are constant and, in a nongrowing economy, the current account is equal to zero. Second, the capital stock is constant and the rate of capacity utilization is at its stationary mean. Real interest rate parity prevails, in the sense that real interest rates are equalized (since the real exchange rate is also in equilibrium, the expected change in the real exchange rate is zero). Finally, there are no changes in reserves or speculative capital movements. The difference between the medium- and long-run NATREX relates to the evolution of net foreign assets and the capital stock. As in the BEER approach, Stein (1999) uses a VECM to empirically implement the NATREX in a single equation context. The misalignment is then calculated as the gap between the estimated value of the VECM and the actual value.

New Open Economy Macroeconomic Class of Models Obstfeld and Rogoff (2001) have demonstrated how the New Open Economy Macroeconomics (NOEM) class of model can be used to calculate the currency movement needed to satisfy internal and external balance (where the latter is taken as a zero current account position). Specifically, it asks the question: How much would the exchange rate have to move to reduce a current account imbalance to zero? There are two key elements in this calculation: knowledge of the consumption elasticity of substitution, which governs how much relative prices have to move in response to changes in the relative consumption of home to foreign goods, and the required movement of the consumption of traded to nontraded goods. If a country has a current

account deficit as a proportion of GDP of 5 percent, for example, balance of payments statistics can be used to calculate how much the relative consumption of home to foreign goods has to move in order to return the ratio to zero. Given the elasticity of substitution, this adjustment can then be used to calculate by how much the exchange rate would have to move in order to facilitate the required consumption switch. The extent of exchange rate adjustment can be calibrated under different scenarios, such as sticky or flexible prices, pricing to market, or the inflation-targeting objectives of the central bank, and using different assumptions regarding the elasticity of substitution.

Since the NOEM approach requires little in the way of data, it would seem to offer a tractable way of calculating how much required exchange rate adjustment is necessary to achieve current account objectives. It may, therefore, be an appealing method of calculating equilibrium exchange rates for developing countries or transition economies where data constraints may make it difficult to implement some of the other approaches referred to earlier.

Other Approaches A somewhat different way of measuring equilibrium exchange rates is to use a time series estimator to decompose a real exchange rate into its permanent and transitory components and to interpret the permanent component as a measure of equilibrium and the distance between the permanent and actual exchange rate as the degree of misalignment. The permanent equilibrium exchange rate has been calculated using a variety of different econometric estimators. This approach often gives time series profiles of the equilibrium exchange rate, which are similar to those created using the BEER and NATREX methods, and is often used as a robustness check on estimates derived from these approaches.

To sum up, there is now a wide range of alternative real approaches to measuring equilibrium exchange rates. Indeed, central banks and practitioners often need to use a combination of these approaches when assessing whether a currency is over- or undervalued.

See also balance of payments; Balassa-Samuelson effect; capital mobility; exchange rate pass-through; exchange rate regimes; exchange rate volatility; interest parity conditions; new open economy macroeconomics; purchasing power parity; real exchange rate

FURTHER READING

Clarida, R., and J. Gali. 1994. "Sources of Real Exchange Rate Fluctuations: How Important Are Nominal Shocks?" *Carnegie Rochester Series on Public Policy* 41: 1 56. Attempts to empirically identify the causes of real exchange rate fluctuations since the collapse of Bretton Woods.

Clark, P. B., and R. MacDonald. 1998. "Exchange Rates and Economic Fundamentals: A Methodological Comparison of BEERs and FEERs." IMF Working Paper 98/67 (March). Washington, DC: International Monetary Fund.

———. 2000. "Filtering the BEER: A Permanent and Transitory Decomposition." IMF Working Paper A0/144. Washington, DC: International Monetary Fund. Two papers that discuss the BEER and FEER approaches to modeling equilibrium exchange rates.

Cumby, R., and J. Huizinga. 1990. "The Predictability of Real Exchange Rate Changes in the Short Run and in the Long Run." NBER Working Paper, No 3468. Cam bridge, MA: National Bureau of Economic Research. Examines whether real exchange rate changes are pre dictable.

Driver, R., and S. Wren Lewis. 1999. *Real Exchange Rates for the Year 2000*. Washington, DC: Institute for In ternational Economics. Monograph that defines what equilibrium real exchange rates are with emphasis on the FEER method and estimates FEERs for the G7 countries for 1995 and 2000.

Faruqee, Hamid. 1995. "Long Run Determinants of the Real Exchange Rate: A Stock Flow Perspective." *IMF Staff Papers* 42 (March): 80 107. Examines the long run determinants of the real exchange rate from a stock flow standpoint.

Faruqee, H., P. Isard, and P. R. Masson. 1998. "A Mac roeconomic Balance Framework for Estimating Equili brium Exchange Rates." In *Equilibrium Exchange Rates*, edited by R. MacDonald and J. Stein. Boston: Kluwer, chapter 4. Discusses and outlines the FEER approach.

Goldstein, M., and M. Khan. 1985. "Income and Price Effects in Foreign Trade." In *Handbook of International*

Economics, edited by R. Jones and P. B. Kenen. Am sterdam: Elsevier, 2:1041 45. Discusses the price and income elasticities of international trade.

Huizinga, J. 1987. "An Empirical Investigation of the Long Run Behavior of Real Exchange Rates." In *Carnegie Rochester Conference Series on Public Policy* 27: 149 214.

Isard, Peter, and Hamid Faruqee, eds. 1998. "Exchange Rate Assessment: Extensions of the Macroeconomic Balance Approach." IMF Occasional Paper No. 167. Washington, DC: International Monetary Fund. Dis cusses and outlines the FEER approach.

MacDonald, R. 1997. "What Determines Real Exchange Rates? The Long and the Short of It." *International Fi nancial Markets, Institutions and Money* 8: 117 53. Presents a reduced form model of the real effective ex change rates of the U.S. dollar, the deutsche mark, and the Japanese yen from 1974 Q1 to 1993 Q2.

. 2007. *The Economics of Exchange Rates: Theories and Evidence*. London: Taylor Francis. Up to date and comprehensive textbook on exchange rate theory and empirics.

MacDonald, R., and P. Swagel. 2000. "Business Cycle In fluences on Exchange Rates: Survey and Evidence." *World Economic Outlook* (*Supporting Studies*): 129 59.

Obstfeld, M., and K. Rogoff. 2001. "Perspectives on OECD Economic Integration: Implications for U.S. Current Account Adjustment." In *Global Economic In tegration: Opportunities and Challenges*. Kansas City, MO: U.S. Federal Reserve Bank, 169 208. Demon strates how the New Open Economy Macroeconomics (NOEM) class of model can be used to calculate the currency movements needed to satisfy internal and ex ternal balance.

Rogoff, K. 1996. "The Purchasing Power Parity Puzzle." *Journal of Economic Literature* 34: 647 68. Compre hensive paper on available evidence (till the mid 1990s) on PPP.

Stein, J. 1994. "The Natural Real Exchange Rate of the United States Dollar and Determinants of Capital Flows." In *Estimating Equilibrium Exchange Rates*, edi ted by J. Williamson. Washington DC: Institute of In ternational Economics, 23 46.

. 1999. "The Evolution of the Real Value of the U.S. Dollar Relative to the G7 Currencies." *Equilibrium Exchange Rates*, edited by R. MacDonald and J. Stein.
Amsterdam: Kluwer Press, chapter 3. Two papers that introduce and apply the NATREX approach to equi librium exchange rate determination.

Williamson, J. 1983. *The Exchange Rate System*. Wash ington, DC: Institute for International Economics.

. 1994. *Estimating Equilibrium Exchange Rates*. Washington, DC: Institute for International Econom ics. Two monographs that introduce and discuss the FEER.

Wren Lewis, S. 1992. "On the Analytical Foundations of the Fundamental Equilibrium Exchange Rate." In *Macroeconomic Modelling of the Long Run*, edited by Colin P. Hargreaves. Aldershot, UK: Edward Elgar, 78 91. Discusses the analytical foundations of the FEER.

RONALD MACDONALD

■ euro

The euro is the currency of the European Monetary Union (EMU). The euro (€) came into existence as an accounting currency on January 1, 1999, for 11 of the then 15 member countries of the EMU (Austria, Belgium, Germany, Finland, France, Ireland, Italy, Luxembourg, Spain, Portugal, and the Netherlands). Greece officially adopted the euro at the beginning of 2001. Britain, Sweden, and Denmark chose not to adopt the euro as their currency, but left the door open to do so in the future. The introduction of the euro marks the first time that a group of sovereign nations voluntarily gave up their individual curren cies in favor of a common currency, and it ranks as one of the most important economic events of the postwar period. Europe created the euro so that it could become a fully integrated economic market like the United States.

The official euro conversion rates for the curren cies of the original 11 members of the EMU were decided on December 31, 1998, by the Council of the European Union based on the recommendation of the European Commission (see table 1). Greece adopted the euro on January 1, 2001, at the con version rate of 340.750 drachmas to the euro. On January 1, 2002, the euro was introduced physically

as a circulating currency, and by February 28, 2002, it became the sole currency of the 12 members of the EMU, also known as the eurozone, euro area, or euroland. From January 1, 1999, until their final withdrawal, the national currencies of the participating countries had been locked against one another at the fixed exchange rates given in table 1.

The euro was introduced in 1999 at the value of $1.18 but, defying almost all predictions that it would appreciate to between $1.25 and $1.30 by the end of the year, it declined almost continuously to a low of $0.82 at the end of October 2000. Starting in February 2002, however, the euro appreciated almost continuously, reaching parity with the dollar in mid-2002 and a high of $1.44 at the end of October 2007. The euro exchange rate with respect to the Japanese yen followed a similar pattern of high volatility.

Euro as Legal Tender The euro is printed and minted, managed, and administered by the European Central Bank (ECB) and the European System of Central Banks (ESCB composed of the central banks of the member states) based in Frankfurt, Germany. The ECB has sole authority for determining the common monetary policy of the EMU.

Table 1
Official Currency Conversion Rates for the Euro (€)

Country	National currency	Currency units per euro
Austria	schilling	13.7603
Belgium	Belgian franc	40.3399
Finland	markka	5.94573
France	French franc	6.55957
Germany	Deutsche mark	1.95583
Ireland	Irish pound	0.787564
Italy	Italian lira	1936.27
Luxembourg	Luxembourg franc	40.3399
Netherlands	guilder	2.20371
Portugal	escudo	200.482
Spain	peseta	166.386

Source: Board of Governors of the Federal Reserve System

Euro banknotes come in €500, €200, €100, €50, €20, €10, and €5 denominations, with the first two seldom used in everyday transactions. Coins come in €2, €1, 50 cents, 20c, 10c, 5c, 2c, and 1c, with the last two denominations seldom used.

Outside the 12 countries of the eurozone, the euro is the legal tender also of the ministates of Monaco, San Marino, and Vatican City (all of which can mint their own coins by agreement with the ECB). Andorra, Montenegro, and Kosovo, as well as eurozone overseas territories of French Guiana, Guadeloupe, Martinique, Mayotte, Réunion, and Saint Pierre et Miquelon also adopted the euro as their currencies, but they do not participate in the ECB or the ESCB.

In May 2004, 10 additional countries joined the European Union (EU), thus increasing its membership to 25 countries. The new members were Poland, Hungary, the Czech Republic, the Slovak Republic, Slovenia, Estonia, Lithuania, Latvia, Malta, and Cyprus. Bulgaria and Romania joined on January 1, 2007, and Albania, Bosnia-Herzegovina, Croatia, Macedonia, Serbia, Montenegro, and Turkey were negotiating accession. The Accession Treaty for the 10 new member states requires them to adopt the euro when they meet some strict criteria, such as a budget deficit not exceeding 3 percent of GDP, a government debt not exceeding 60 percent of GDP, a low inflation rate, an interest rate close to the EU average, and a fairly stable exchange rate. Slovenia adopted the euro on January 1, 2007, but the other new members were not expected to adopt the euro for several more years.

Euro as an International Currency From its birth, the euro was destined to become an important international currency because the EU (1) is as large an economic and trading unit as the United States, (2) has a large, well-developed, and growing financial market, which is increasingly free of controls, and (3) is expected to have a good inflation performance that will keep the value of the euro stable. If the international use of the euro were to match the EU's share of world GDP, exports, and financial market, the euro would become as important as the dollar as an international currency. This would mean that the

relative international use of the dollar would fall to 40 45 percent of the total (from 55 60 percent in 2007), with an equal share going to the euro and the remainder going mostly to the yen and a few other smaller currencies, such as the Swiss franc, the Canadian dollar, and the Australian dollar but mostly the yen.

There are equally good reasons why the euro probably will not displace the U.S. dollar as the leading international currency in the short term. These are: (1) most primary commodities (especially petroleum) are priced in dollars, (2) most non-European countries are likely to continue to use the dollar for most of their international transactions for the foreseeable future, with the exception the former French colonies in West and Central Africa; and (3) sheer inertia, which favors the incumbent (the U.S. dollar).

See also common currency; dollar standard; dominant currency; European Central Bank; European Monetary Union; Maastricht Treaty; optimum currency area (OCA) theory; reserve currency; vehicle currency

FURTHER READING

Board of Governors of the Federal Reserve System. 1999. "The Launch of the Euro." *Federal Reserve Bulletin* (October): 655 66. Examines the process of launching the euro.

Issing, Otmar. 2005. "The ECB and the Euro." *Journal of Policy Modeling* (June): 405 20. Examines the operation and performance of the ECB since its creation.

Mundell, Robert A. 1961. "A Theory of Optimum Cur rency Areas." *American Economic Review* (November): 509 17. Analyzes the benefits and costs of adopting a single currency, such as the euro.

———. 2000. "The Euro and the Stability of the Inter national Monetary System." In *The Euro as a Stabilizer in the International Economic System*, edited by Robert A. Mundell and Armand Clesse. Norwell, MA: Kluwer, 56 84. Examines the effect of the introduction of the euro on the dollar and on the functioning of the inter national monetary system.

Salvatore, Dominick. 1997. "The Unresolved Problems of the EMS and EMU." *American Economic Review* (May): 224 26. Examines the problems in the operation

of the EMU and its precursor, the European Monetary System.

———. 2002. "The Euro: Expectations and Performance." *Eastern Economic Journal* (winter): 121 36. Evaluates the performance of the euro since its creation.

DOMINICK SALVATORE

■ Eurocurrencies

Eurocurrencies are domestic currencies of one country on deposit in a second country. More precisely, the term *Eurocurrency* means the currency denomination of a bank deposit or loan that is on the books of a bank situated outside the currency's country of origin. For example, large amounts of U.S. dollar deposits and loans are booked in banks located in London, Frankfurt, Tokyo, Singapore, and many other international financial centers. Eurocurrencies are on the books of major banks with international operations, but not necessarily banks headquartered in the Eurocurrency's country of origin. For example, Eurodollar deposits and loans are offered by banks of many home jurisdictions, not just by offshore branches and subsidiaries of American banks.

More than 70 percent of the world's Euro-currencies are in U.S. dollars, with the rest in euros, yen, British pounds, Canadian dollars, Swiss francs, and Australian dollars. Individually they are often called Eurodollars, Euroeuros, Euroyen, and so on. The prefix *Euro* is an anachronism, since large amounts are booked in banks outside Europe. For a time, dollar deposits in Japan were called "Juro-dollars," Singapore "Singdollars," and so on, but the generic term *offshore* for these deposits and loans has become more common.

Eurodollar deposits were first offered in the 1950s by London- and Paris-based banks (particularly the Moscow Narodny Bank in London and the Banque Commerciale pour l'Europe du Nord in Paris) in order to accommodate the desire of the Soviet Union and other communist countries to avoid holding their large dollar surpluses in the United States. But the Eurodollar market first expanded to international

significance with the incursion of U.S. corporations into Europe in the 1960s. An added impetus was the imposition in the mid-1960s of a ban on U.S. corporations' borrowing in the United States for operations abroad, reflecting concern about a growing balance of payments deficit. U.S. firms quickly discovered that they could circumvent these regulations by borrowing in dollars from London-based banks. What began mostly in London with Eurodollars soon spread to the world's major financial markets and to other major currencies.

Causes of Growth The long-term growth of the Eurocurrency market was driven neither by politics nor by regulatory worries, but instead by economics. From the beginning, the British authorities set a precedent by not attempting to regulate Eurocurrency banking. To this day, in all of the world's major financial centers, neither home- nor host-country regulators impose any of the costly burdens that they place on domestic banks: requirements such as holding noninterest-bearing reserves, interest rate controls, deposit insurance, or high taxes. Typically also, restrictions on the entry of new banks are minimal. Moreover, "Eurobanks" that is, the international departments of big banks that offer offshore deposits accept only wholesale deposits, mostly $1 million or more. As a result of all these regulatory and economic advantages, Eurobanks are able to offer interest rates on deposits that are higher than in the currency's home country, and also interest rates on loans that are lower. In short, the "spreads" between loan and deposit rates are lower on Eurobanks' offshore deposits than at home.

A major impetus for growth of Eurobanking came with the virtual doubling of oil prices in the fall of 1973. In the context of heightened tension with the United States over its support for Israel, Middle Eastern oil exporters deposited much of their expanding oil revenues in Eurodollars in London. At the same time, oil-importing countries were hungry for hard currency loans. The Eurocurrency market expanded rapidly to meet this demand. Doomsayers had predicted disaster if the oil revenues were not "recycled." The International Monetary Fund (IMF) had until then been the conventional lender to deficit countries but lacked capacity to recycle such unprecedented amounts. The Eurocurrency markets grew seamlessly and saved the day.

By the late 1970s, the value of Eurodollar deposits exceeded the value of dollar deposits within the U.S. banking system. In December 1981, U.S. banking legislation was revised to allow U.S. banks to offer "international banking facilities" (IBFs) that is, dollar deposits and loans to nonresidents of the United States and to other IBFs that are immune from U.S. regulatory requirements. Hence U.S. banks could offer deposit and loan rates competitive with the larger Eurodollar markets in London and elsewhere. More than 75 percent of U.S. IBF banking is in New York, with almost all the rest in California and Illinois.

Interest Rates The Eurocurrency markets virtually pioneered the offering of long-term loans with floating interest rates. They "unbundled" the usual package, which fixed loan rates for the term of the loan. Eurocurrency deposits typically mature every 30 to 90 days; so to induce rollover of existing deposits, or to attract new deposits, current deposit rates must be paid.

Eurocurrency loan rates are usually revised every three to six months, based on current deposit rates. This protects banks from the interest rate risks of long-term maturity mismatching. More precisely, loan rates (in London) are based on the London Inter Bank Offered Rate (LIBOR), which is reset daily at the median rate that top-tier London-based Eurobanks offer to pay other banks for short-term deposits. Other offshore financial centers quote rates close to LIBOR. LIBOR is usually slightly above the rate paid on deposits from nonbank customers since interbank deposits can be acquired on demand, within minutes. In effect, Eurocurrency loans are based on banks' marginal cost of raising new funds. The actual rates charged are LIBOR + , where the premium varies from borrower to borrower depending on credit risk. Very creditworthy customers can borrow at LIBOR or even, occasionally, slightly below. In this sense, LIBOR is like the "prime" rate in the United States and other domestic markets, with the major difference that it is marked to market daily. Much of the Eurodeposit market is securitized: they

are certificates of deposit that are traded continuously; hence their yield varies continuously.

The practice of making long-term loans with variable rates was a key trigger for the so-called international banking crisis of 1982–89. In early 1980, the newly installed chairman of the U.S. Federal Reserve System, Paul Volcker, declared war on inflation, which had risen to double-digit levels in the United States and elsewhere. He more than doubled short-term interest rates, from about 10 percent to more than 20 percent. Because of the tight linkages to Europe and the rest of the world via competition between New York and London for Eurodollar deposits, short-term rates, based on LIBOR, doubled worldwide. By 1982, middle-income countries in East Asia, Eastern Europe, and especially Latin America had borrowed heavily via syndicated loan packages put together in the Eurocurrency markets. They were paying LIBOR + on their loans. By late 1981, their debt service burdens had doubled. This triggered a series of severe slowdowns and partial defaults, beginning with Mexico in August 1982.

Size of Market The size of the offshore markets including interbank loans as of 2007 was roughly $2 trillion in London, $1.5 trillion in New York, $1 trillion in Germany, and more than $0.5 trillion each in the Cayman Islands, Switzerland, and Japan. Net of interbank loans, the numbers are only about one-quarter of these. In other words, London Eurobanks had roughly $0.5 trillion worth of loans to nonbanks on their books in 2007. Nevertheless, interbank loans are important because they constitute the core of the world's trading in foreign currencies. Foreign exchange trading is part and parcel of the enormous offshore interbank market because Eurobanks typically find themselves with more deposits than loans in any particular currency, or vice versa. Also, unlike domestic retail banks, they often negotiate loan agreements before obtaining funding for them. For both reasons, Eurobanks trade trillions of dollars of deposits daily.

For example, a Japanese bank in London may find itself with $2 billion worth of yen deposits but only $1 billion worth of yen-denominated loans, with another $1 billion denominated in dollars. To avoid currency risk, the bank might "buy"—that is, borrow—$1 billion in dollar deposits (offering the going LIBOR interest rate for dollars) and "sell"—that is, lend—$1 billion in yen deposits. The transaction often takes the form of a "swap" of yen for dollars that is scheduled to be reversed in a few weeks or months when the $1 billion loan matures. In either case—whether it is structured as an outright sale of yen and purchase of dollars, or a swap—$1 billion of foreign exchange has been traded.

As another example, a London-based German bank that has already more or less matched the currency denominations of its loans and deposits may negotiate a $2 billion euro loan. In order to fund the loan it must buy deposits, offering the going LIBOR rate for euros on the interbank market. Because the largest volume of offshore interbank trading takes place in London, LIBOR is the international benchmark for short-term lending, with modifications for currency denomination, term to maturity, and risk.

In short, the Eurocurrency, or "offshore" interbank market for deposits, is the main means by which almost $2 billion of foreign exchange is traded daily. More fundamentally, the Eurocurrency market—which is in fact not just European but worldwide—is the main means by which the demands for and supplies of loanable bank funds are matched up internationally.

See also capital flows to developing countries; dominant currency; euro; International Monetary Fund (IMF); money supply; offshore financial centers; petrodollars, recycling of; reserve currency; vehicle currency

FURTHER READING

Dufey, Gunter, and Ian Giddy. 1994. *The International Money Market*. 2d ed. Upper Saddle River, NJ: Prentice Hall. An updated edition of a 1978 textbook that was a pioneering and eminently clear exposition of the evolving Eurocurrency markets, and much more.

Husted, Steven, and Michael Melvin. 2007. *International Economics*. 7th ed. Boston: Pearson Addison Wesley. Contains one of the best treatments of Eurocurrencies from a macroeconomic standpoint.

Moffett, Michael, and Arthur Stonehill. 2006. *Funda mentals of Multinational Finance*. 2d ed. Boston: Pearson Addison Wesley. Contains a good primer on the origins, institutions, and workings of Eurocurrency markets.

Rivera Batiz, Francisco L., and Luis Rivera Batiz. 1994. *International Finance and Open Economy Macro economics*. Upper Saddle River, NJ: Prentice Hall. Theoretically sophisticated, clarifies links between the microeconomics of Eurocurrency markets and the macroeconomics of open economies.

JAMES W. DEAN

■ European Central Bank

The European Central Bank (ECB) is the central bank for the euro area. It opened its doors in June 1998 in Frankfurt, Germany. Together, the ECB and the national central banks that are members of European Monetary Union (EMU) form the European System of Central Banks (ESCB), also known as the eurosystem.

Before the ECB began to take responsibility for the conduct of monetary policy in the euro area, the member countries created the European Monetary Institute (EMI) in 1994 to facilitate the work toward transition to the era of the single currency. Its primary task was to address the technical needs for the switch to the euro. The EMI was responsible for reporting to the European Union (EU) on the progress toward convergence by individual member states, as pre-scribed by the Maastricht Treaty. The conduct of monetary policy remained under the responsibility of the individual member states. By the end of 1995, member governments approved a plan for change that focused on the need for rigorous acceptance of the convergence requirements. A legal basis for the single currency was necessary so that all who used the euro could be confident that it would be an accepted means of payment.

Objectives and Tasks of the ECB The primary objective of the ECB and, by implication, the ESCB, is to maintain price stability, a goal set but not de-fined by the Maastricht Treaty. The ECB defines price stability as an inflation rate close to 2 percent in

a Europewide index of consumer prices called the Harmonized Index of Consumer Prices over the "medium term." What is the medium term? The ECB does not provide a precise definition. It notes that monetary policy acts with long and variable lags, an idea made famous by Milton Friedman. Experi-ence, however, suggests that inflation can be con-trolled within the objectives set out by the ECB within a two-year horizon. To achieve this objective, the ECB decided to give prominence to money growth as well as to a wide range of indicators such as the exchange rate, the yield curve, and various fiscal indicators. More precisely, the ECB's strategy involves a so-called two-pillars approach. This means that, in setting interest rates for the euro area, the ECB considers developments both on the real side of the economy, which are of a shorter-term nature, as well as on the monetary side, which cap-tures the longer-run influences on monetary policy. Since current interest rate decisions have an impact on future decisions by individuals and firms, the ECB also receives guidance from forecasts of infla-tion and real economic activity for the euro area as a whole.

The ECB is headed by a president who serves a nonrenewable eight-year term. The first president of the ECB, Wim Duisenberg from the Netherlands, did not serve his full term, which would have ended in 2006; instead, Jean Trichet of France was ap-pointed as the president of the ECB in 2003. Since the president and other senior officials of the ECB are appointed by "common accord" by the heads of state or governments of EU members, national political imperatives play an important role in such appoint-ments. Therefore, it is unlikely that a country could have more than one representative on the executive board.

The principal decision-making body of the ECB is the governing council made up of the governors of the thirteen euro-area national central banks and the executive board. The executive board, which consists of six members chosen from the governing council, is responsible for the implementation of monetary policy and carrying out the day-to-day affairs of the ECB. The governing council of the ECB consists of

members of the executive board and all the heads of the national central banks that belong to the euro area. The governing council is mainly responsible for formulating monetary policy. The general council, largely an advisory body, includes members from both euro-area and non-euro-area countries that are members of the EU.

Both the executive board and the governing council meet twice a month, generally in Frankfurt. During the first monthly meeting the governing council announces the monetary policy decision made by the executive board. The second monthly meeting is reserved for making decisions related to the other tasks of the euro system. The schedule of meetings is published in advance so financial markets can prepare for the announcement of the interest rate decision. Additionally, in an emergency or crisis the governing council can meet in an extraordinary session, as happened, for example, following the terrorist attack on the United States in September 2001. The president of the ECB announces the interest rate decision immediately after the meeting and holds a press conference, but the ECB does not release minutes of the meeting.

In addition to its responsibility for defining and implementing monetary policy in the euro area, the ECB conducts foreign exchange operations, holds and manages member states' foreign exchange reserves, and helps promote the Europewide payments system. This payment system is called TARGET (Trans-European Automated Real-Time Gross Settlement Transfer System). Finally, the ECB, together with the individual national central banks, collects and disseminates a large variety of financial and economic statistics.

The primary monetary policy instruments consist of open market operations followed by a marginal lending facility that permits select financial market participants to borrow overnight from national central banks against eligible assets. Finally, banks in the EMU are required to hold reserves against short-term deposits.

Enlargement In 2004, a historic enlargement took place. Ten new member states joined the EU on May 1: the Czech Republic, Estonia, Cyprus, Latvia, Lithuania, Hungary, Malta, Poland, Slovakia, and Slovenia. In 2007, Romania and Bulgaria joined the EU. All of these countries must fulfill the Maastricht Treaty convergence requirements. The enlargement of the EU in 2004 introduced a potential organizational problem for the ECB's governing council. With 15 national central banks and six executive board members, the main decision-making body of the ECB originally consisted of 21 members. Following enlargement, membership would rise to a total of 31 members and, potentially, a still larger number once other countries are admitted into the EU. Therefore, in 2002, the ECB recommended a voting system that would limit the number of voting national central bank members to a maximum of 15. This means that there would be a rotation system among members of the eurosystem. The rotation system is designed to ensure that the national central bank governors with the right to vote are from member states that, taken together, are representative of the euro-area economy as a whole.

In spite of the growing pains experienced by the ECB, it remains a grand experiment that shows every sign of success, durability, and future expansion. The euro, which is the most visible expression of the ECB's existence, is a currency that may someday rival the U.S. dollar as a reserve currency. Finally, the economic size of the euro area is now larger than that of the U.S. economy, and it is likely that economic performance in Europe will have repercussions worldwide.

See also common currency; euro; European Monetary Union; Federal Reserve Board; Maastricht Treaty; optimum currency area (OCA) theory

FURTHER READING

de Haan, Jakob, Sylvester Eijffinger, and Sandra Waller. 2005. *The European Central Bank*. Cambridge, MA: MIT Press. Provides a comprehensive historical analysis of the ECB, analyzes its policy functions, and provides an early general assessment of its performance.

European Commission. 1999. *Third Stage: Organization and Implementation* (*from 1 January 1999*). Downloadable from http://europa.eu/scadplus/leg/en/s01030.htm. One of many official documents that provide the legal

text defining the principal economic requirements for monetary union.

Padoa Schioppa, Tommaso. 2004. *The Euro and Its Central Bank: Getting United after the Union*. Cambridge, MA: MIT Press. Written by an insider, a former governor at the ECB and Italy's finance minister from 2006 to 2008, who traces the creation of the ECB and its role in the project to create more integration in Europe.

Scheller, Hanspeter K. 2007. *The European Central Bank: History, Role and Functions*. Rev. ed. Frankfurt: European Central Bank. Downloadable from http://www.ecb.int/ecb/html/index.en.html. An official publication that provides background and other details about the structure, organization, and functions of the ECB.

PIERRE L. SIKLOS

■ European Monetary Union

The establishment of the European Monetary Union (EMU) was an event without historical precedent. Although monetary unions have existed in the past, the EMU was the first union of sovereign states to agree on a single currency and a single monetary policy, but not on a common fiscal policy. The development of the EMU was an evolutionary process. The drive for greater European integration began shortly after World War II, but the actual path to the EMU originated with the European Monetary System (EMS), an initiative led by France and Germany in 1978 to stabilize exchange rates within Europe following the end of the postwar Bretton Woods exchange rate system.

Even before the Bretton Woods system of exchange rates collapsed during the 1970s, six European countries Belgium, France, West Germany (as it was then called), Italy, Luxembourg, and the Netherlands signed the Treaty of Rome in 1957, which led to the creation of the European Community (EC). This was an ambitious attempt, first to integrate the economies in Europe, and then to use such integration to fix currencies among members of the EC. Although the EMS fixed exchange rates between the member states, the arrangement permitted

bilateral exchange rates to fluctuate within a particular range or target zone. In practice, however, the rates were fixed to the German currency, the deutsche mark. When market-determined exchange rates came close to the permissible bands, central banks in the system intervened by buying or selling foreign exchange, usually German deutsche marks, to push the exchange rate back into the band.

Although most countries in the world participated in the Bretton Woods pegged exchange rate arrangement, almost immediately after the creation of the EMS, the original six members agreed that permissible exchange rate fluctuations would be limited to a band, or target range, narrower than that permitted by the International Monetary Fund at the time. The International Monetary Fund is one of two institutions, the other being what was eventually to be called the World Bank, that would govern the postwar international exchange rate system. The new European arrangement would be called the "snake within the IMF tunnel." In other words, the EMS was a miniature version of the Bretton Woods arrangement. The "snake," however, never functioned smoothly. (West) Germany and the Netherlands, for example, allowed their currencies to float temporarily. Eventually this system was abandoned in 1973, when a devaluation of the U.S. dollar led the EC to float all of its currencies for a time. Nevertheless, EC members reaffirmed their commitment to a fixed exchange rate system as a precursor to achieving eventual monetary and economic union.

Exchange Rate Mechanism In 1974, the second stage of the plan that would eventually lead to a currency union began when EC members pooled their foreign exchange reserves, increased short-term credit facilities to member states, and intensified coordination of members' economic and monetary policies. Despite a variety of setbacks and delays, the new and improved version of the EMS started operation in March 1979. Initially there were only six members in this club, namely the signatories of the original Treaty of Rome. Later, membership increased: Italy and Ireland joined in 1978, Greece in 1981, Spain and Portugal in 1986. The last holdout was the United Kingdom, which joined the

Exchange Rate Mechanism (ERM) in 1990, although it withdrew, along with Italy, in September 1992. Meanwhile, even nonmember countries, such as Sweden and later Denmark, effectively belonged to the EMS system by adhering to its rules, albeit not on a formal basis.

The system's monetary unit would be called the ECU or European Currency Unit. The ECU represented a basket of currencies weighted roughly by the relative importance of member countries' economies. The ECU would serve as a kind of *numéraire*, or unit of account, for the ERM, as there were no circulating ECU notes or coins. The EMS stipulated allowable movements of ±2.25 percent (±6 percent for Spain) in either direction of one another's currencies values, and an additional mechanism was added to provide economic support to the relatively less prosperous members of the EC. Allowable fluctuations were relative to a central parity, essentially the midpoint of the band around which the exchange rate was permitted to fluctuate. The EMS did permit an exchange rate realignment (revaluation) of the currencies that could not be sustained within the zone, but this was supposed to be only a last resort. For example, if one country experienced severe balance of payments difficulties, then by common agreement there would be a call for a realignment among the EMS members.

In spite of an agreement to fix exchange rates rigorously, there were 27 realignments of the currency relationships between 1979 and 1983, and 12 more during the 1984 87 period. After that, however, no further realignments were made, until September 16, 1992, known as "Black Wednesday," when the United Kingdom and Italy left the EMS. Also, Ireland, Spain, and Portugal all devalued their currencies against the German currency, the deutsche mark. The European exchange rate crisis of 1992 threatened to derail the whole EMU project. Instead, in August 1993, the bands were widened to ±15 percent, except for the Netherlands, which retained the narrower fluctuation band. This may have had the effect of reducing the incentive to test the resolve of the potential members of the single currency area to carry on with their project, although

economic recovery in Europe, assisted by an expansionary fiscal policy in Germany following reunification, also facilitated the return to a course toward the creation of the euro.

In principle, the chief advantage of the target zone system (ERM) is its ability to prevent large swings in a country's current account when there are sharp swings in the nominal exchange rate. Of course, the main drawback with EMS-type systems is that they may also be subject to speculative attacks. After all, the limits set by policymakers on exchange rate fluctuations will hold only if the policies of the countries that belong to the target zone system are credible and compatible with one another. Indeed, illustrations that advocates of floating rates point to, in favoring a more flexible exchange rate systems, are the impact of the large devaluation of the Mexican peso in early 1995 and the devaluations in several Asian countries in 1997. Advocates of pegged exchange rates reply by drawing attention to the sharp appreciation of the Japanese yen in 1992, the rapid appreciations of the Australian, New Zealand, and Canadian dollars in 2005 6, as well as the large swings in the euro U.S. dollar exchange rate since its introduction in 2002.

The EMS continues to exist as ERM II, which took effect on January 1, 1999, and resembles the original ERM, except that the fluctuation band of 15 percent became the accepted norm. The countries that joined the European Union (EU) must adhere to its rules prior to being admitted into the EMU and must belong to ERM II for at least two years.

Although the EMS and ERM were intended to foster greater economic integration within Europe, the actual blueprint for the single currency area can first be traced to the Werner Report of 1970. This report followed on the heels of a summit of European heads of state and government in 1969 held at the Hague. This meeting led to a decision that, henceforth, the EMU ought to be an explicit goal for the EC. Werner was the prime minister of Luxembourg, who was entrusted to provide a way to create a common currency area. The Werner Report proposed a reduction in the size of the fluctuation bands

for the currencies in the EMS in a first stage, a removal of all barriers to financial flows and enhancement of the integration of financial and banking sectors in a second stage, and an irrevocable fix in exchange rates in a third stage. Interestingly, the report was agnostic about the need for a single currency, as opposed to several currencies with a fixed exchange rate. Nevertheless, it called for greater fiscal harmony among the member states if the goal of a monetary union of some kind were to succeed. As mentioned earlier, the collapse of the Bretton Woods system interrupted the drive to the EMU, and problems with the EMS did not help matters either.

Delors Report and Beyond The next attempt to lay out a road map to monetary union was the Delors Report of 1989, named after the former French politician who is also a former president of the European Commission. Critical to achieving monetary union were the so-called convergence criteria. The goal, in principle, was simple: to ensure that the members of the area that would form a monetary union would have similar macroeconomic environments, as this would facilitate the introduction of a fixed exchange rate among the member currencies that would not be altered, which would pave the way for the introduction of a common currency.

One of the innovations of the Delors Report was its outline of the institutional environment under which monetary union would operate. A European System of Central Banks (ESCB) was to be formed and this body would formulate and implement monetary and exchange rate policy in Europe. The Maastricht Treaty would later enshrine the convergence criteria that were believed to be necessary for a successful monetary union, and it contained a recommendation to create a European Central Bank (ECB) that would bring together the ESCB under a separate institutional umbrella.

The need for economies that participate in a monetary union to look alike, in broad economic terms, is associated with the work of Robert Mundell, who was awarded a Nobel Prize for his work demonstrating the conditions that countries would have to fulfill to form an optimum currency area. In spite of the intellectual debt to Mundell's work, the EMU

project was fundamentally a political one, namely to encourage and enshrine integration among European countries, albeit underscored by the economic benefits of a single currency. These benefits are thought to include a reduction in transactions costs. Although tourists no doubt have seen the advantages of not having to exchange one currency for another as frequently, the evidence that the reduction of transactions costs is of macroeconomic significance is more dubious. As for the single currency fostering greater integration, the jury is still out, although there is evidence that participation in a currency area, or in a monetary union, can contribute to greater similarities in business cycle performance among the participants.

The establishment of the ECB was preceded by the European Monetary Institute (EMI). The creation of this institution marks the beginning of stage II toward the EMU. The EMI's task was to prepare the groundwork, at both the policy and the institutional level, until the ECB took over, that is, until the irrevocable exchange rates became a reality.

A second innovation of the Delors Report was the proposal for a single currency that would follow a transitional period leading up to the setting of irrevocable exchange rates. Unlike the Werner Report, however, there was considerably less emphasis on fiscal harmony within the proposed monetary union. The thinking was that the convergence requirements would represent sufficient discipline to bring the countries into line with one another. Events were to prove otherwise, as the Maastricht Treaty was later accompanied by the Stability and Growth Pact (SGP). The SGP was intended to generate a form of fiscal discipline that resembled the monetary discipline required of the Maastricht Treaty.

The original Delors Report called for European Monetary Union (EMU) by 1999. At the end of 1995, leaders of the EU announced in Madrid plans for a single currency that would be called the euro, and they reaffirmed the launch date of January 1, 1999. Events would delay the introduction of the euro but, in spite of considerable skepticism, the single currency project eventually saw the light of day. The goals of monetary policy for the euro area,

and the conversion of accounting standards, were agreed to and implemented by 1999. Eleven EU member states initially became members of EMU in May 1998, marking the start of the third and final stage toward the EMU, which effectively came into being on January 2, 2002. Hence, between January 1, 1999, and December 31, 2001, the euro replaced the ECU at an exchange rate of one to one, while EMU member states' exchange rates were irrevocably fixed at rates announced on December 31, 1998.

See also band, basket, and crawl (BBC); Bretton Woods system; common currency; currency crisis; discipline; euro; European Central Bank; exchange rate regimes; fear of floating; international reserves; Maastricht Treaty; optimum currency area (OCA) theory

FURTHER READING

Bayoumi, Tamim, and Barry Eichengreen. 1993. "Shocking Aspects of European Monetary Integration." In *Adjustment and Growth in the European Monetary Union*, edited by Francisco Torres and Francesco Giavazzi. Cambridge: Cambridge University Press, 193 229. A somewhat technical study that asks whether greater monetary integration might be reflected in the pattern of economic shocks that hit prospective members of the monetary union.

De Grauwe, Paul. 1997. *The Economics of Monetary Integration*. Oxford: Oxford University Press. Covers both the economic and political aspects of the EMU project.

Delors Report, Committee for the Study of Economic and Monetary Union. 1989. *Report on Economic and Monetary Union in the European Community*. Luxembourg: Office for Official Publication of the European Communities. Text of the report that outlines a strategy for monetary integration, written in somewhat legalistic terms.

Eichengreen, Barry. 1993. "European Monetary Unification." *Journal of Economic Literature* 31 (September): 1321 57. A broad survey of the literature and the pros and cons of European monetary integration. Note that the publication date precedes actual monetary union in Europe.

Gros, Daniel, and Niels Thygesen. 1992. *European Monetary Integration*. New York: Longman. Like the De Grauwe volume, covers the essentials on the development of ideas and policies that would foster the dream of monetary union in Europe.

Padoa Schioppa, Tommaso. 2000. *The Road to Monetary Union in Europe: The Emperor, the Kings, and the Genies.* Oxford: Oxford University Press. The story of the EMU written by an insider, a former member of the governing council of the ECB, and later, minister of finance for Italy.

PIERRE L. SIKLOS

■ European Union

The European Union is both a vague aspiration and a concrete organization. The aspiration, never held by more than a small minority of Europeans, is for a federation of European states to replace the hodgepodge of sovereign states that has characterized the continent in modern times. The organization European Union (EU) is a complex entity with federal features that is less than a full-fledged state yet far more than a traditional international grouping of states. After it acquired two new members in January 2007, the EU comprised 27 countries with a combined population of nearly 500 million and a combined gross national product of approximately €10 trillion. Perhaps it is best understood as an association of countries (member states) that have agreed to share sovereignty in a number of policy areas and to coordinate closely in others, in order to pursue jointly key objectives such as economic growth, social protection, regional stability, and global security.

The EU has grown in size and stature over the years but is increasingly beset by serious political problems. Buffeted by weak public support, poor leadership, the myriad challenges of enlargement, and widespread concerns about the impact of globalization on Europe's generous welfare states, the EU faces an uncertain future. Although hardly destined to wither away or collapse in ignominy, neither is the EU likely to evolve into a superstate, let alone a superpower, as some ardent advocates of European integration would like to see happen.

Origins Today's EU originated in the European Coal and Steel Community, an organization that, in our age of bits and bytes, sounds quaint or even archaic. In the early 1950s, however, coal and steel were essential ingredients for economic recovery and resurgence, and cherished symbols of industrial strength and vitality. The name of the organization denoted the political and economic importance of coal and steel in postwar Europe as well as the special nature of the relationship among its member states. Despite an apparently narrow emphasis on industrial restructuring and revival, the new entity was a "community" and not simply a functional international organization. As such, its members acknowledged the necessity of reconciling national interests with common, overarching security concerns in order to avoid a repetition of Europe's disastrous past and help ensure a more peaceful and prosperous future. By launching the Coal and Steel Community, the original member states also launched the process of political and economic integration that has continued in Europe, through various ups and downs, to the present day.

There was nothing inevitable about the form or content of the EU's precursor. Initial discussions about European unity in the postwar period focused on grandiose federal schemes. In a famous speech in Zurich in 1946, Winston Churchill, Europe's best-known statesman, endorsed the idea of a United States of Europe. Altiero Spinelli, a veteran Italian antifascist and ardent European federalist, advocated a "big-bang" approach to sharing national sovereignty. War weariness and popular disillusionment with the status quo seemed propitious for bold international initiatives. Yet when Europe's leaders met in The Hague in 1948 to discuss political integration, rhetoric clashed with reality. Few of them were willing to surrender sacred national sovereignty to an all-encompassing European federation. What emerged instead was the Council of Europe, an arena for exchanging ideas and information on European integration that lacked political clout.

Meanwhile, Europeans were grappling with the challenges of reconstruction. The Marshall Plan had provided the means, in the form of scarce American dollars, to fuel Western Europe's rapid economic recovery. Motivated in part by escalating East-West tensions, it also accelerated the onset of the Cold War. In view of rapid economic recovery and deepening Cold War hostility, the fate of Germany became more pressing than ever. In 1948 the Western Allies agreed to establish the Federal Republic of Germany. Nevertheless, France did not want the new West German state to become economically powerful, at least not yet. In particular, France coveted Germany's abundant coking coal and sought to modernize its own steel industry before German mills and market share could fully recover from the destruction of the war. Such a policy was unacceptable to the United States, which sought to reconstitute Germany politically and economically as the cornerstone of Western Europe's resurgence and a bulwark against the spread of communism. Under intense American pressure, France had little choice but to come up with a new approach toward its historical enemy.

In response, Jean Monnet, a senior French civil servant and confidant of leading American politicians, proposed the Coal and Steel Community. Rather than pursue a beggar-thy-neighbor policy toward each other, France and Germany would cede responsibility for managing the coal and steel sectors to a supranational High Authority. Robert Schuman, foreign minister of France, announced the proposal in May 1950 and pushed it through a skeptical French government. German chancellor Konrad Adenauer strongly supported what became known as the Schuman Plan, which provided a means of binding the new Federal Republic into Western Europe, promoting reconciliation with France, and lifting Allied controls on Germany's heavy industry.

In his public declaration, Schuman said that the plan was open to the other countries of Europe. With the onset of the Cold War, in effect that meant Western Europe. Britain, with its newly nationalized industries, different war experience, and unwillingness to share sovereignty, decided not to participate. A number of other countries followed suit, whereas Spain, an international pariah after the war, was excluded by default. That left only Italy, which saw

European integration largely in anticommunist terms, and the three Benelux countries (Belgium, Netherlands, and Luxembourg), which were tied economically to Germany, to take up Schuman's offer. As a result, the Coal and Steel Community came into existence in 1952 with six member states.

Monnet had wanted the institutional structure of the new community to comprise only the High Authority and a Court of Justice. In the event, governments successfully pressed for the inclusion as well of a Council of Ministers, to provide a direct national input into decision making, and an Assembly, to lend democratic credence to the organization. Thus the institutional design of today's EU took shape, with the High Authority changing its name to the European Commission in the late 1950s, following the establishment of the European Economic Community (EEC).

The Coal and Steel Community played an unglamorous yet essential role in the postwar European settlement. Its significance was more political than economic, notably by providing a diplomatic solution to the contentious question of German reconstruction and rehabilitation. Conversely, the emergence of the EEC later in the decade had less to do with Franco-German rapprochement than with the immediate economic interests of the contracting states. In particular, the explosive growth of international trade in the 1950s led the Dutch, traditional advocates of economic liberalism, to push for a Western European common market. Adenauer, enamored of any idea that would strengthen European integration, took up the call, despite the objections of Ludwig Erhard, his economics minister, who feared that the proposal would damage prospects for global trade liberalization. France was torn between accepting greater competition, albeit in a regional framework, and maintaining a policy of trade protectionism. Champions of the common market in the French government won the day by including in the proposed community provisions for EEC-level agricultural subsidies and a preferential regime for current and former colonies of the member states. With a huge, heavily subsidized agricultural sector

and a large, fragmenting empire, France stood to gain handsomely from such side payments. Hence the origin of the Common Agricultural Policy (CAP) and the Cotonou Agreement, two of the most controversial, trade-distorting policies of the EU.

The same six members of the Coal and Steel Community went on to establish the broader Economic Community in 1958, Britain again having refused to participate. The launch of the EEC, later known simply as the European Community (EC), was far from an earth-shattering event. Few Europeans had followed the negotiations, which were highly technical and conducted mostly in secret. News at the time was dominated by Soviet threats against Berlin, the deteriorating situation in Algeria, and the imminent collapse of the French Fourth Republic. The EC, which would later insinuate itself into everyday life in Europe and, in the form of the EU, occasionally dominate domestic politics, began with a whimper, not a bang.

Development of the EC Almost from the beginning, the EC developed along two lines: widening and deepening. Widening refers to enlargement, as more and more countries requested membership. Deepening refers to the acquisition by the EC of responsibility for or involvement in additional policy areas, as member states agreed to transfer more sovereignty to the organization. Conceptually, widening and deepening are contradictory: the larger the membership, the more cumbersome and less cohesive the organization. The EU may have reached that point by 2007, although even with 27 members the institutions operated surprisingly well. Historically, in any case, widening and deepening went hand in hand. Indeed, member states deliberately undertook institutional reform and increased the EC's policy scope in part to ensure that wider would not mean weaker.

The first opportunity to do so came almost immediately after the establishment of the EC. Having first dismissed European integration as unworthy of a victorious Great Power, and then having tried to subvert the EC by establishing the rival European Free Trade Association, Britain accepted the reality of growing economic interdependence with the

Continent and applied for membership in 1961. It would be another 12 years before Britain   together with Denmark and Ireland   joined, thanks largely to the policy concerns and personal spite of French president Charles de Gaulle. De Gaulle wanted to ensure that the EC was ready for Britain, meaning that the CAP, and with it the EC's budget, were fully fleshed out. Because of different agricultural preferences, Britain would have objected to the establishment of the CAP in the form in which it eventually emerged in the late 1960s. By the time Britain joined, the CAP was already in place. By contrast, Britain subscribed fully to the customs union, which finally came into being in 1968 following the eradication of tariffs and abolition of quotas among member states. As a corollary to the customs union, the EC developed a common commercial policy, the basis for collective member state action in successive rounds of multilateral trade liberalization under the auspices of the General Agreement on Tariffs and Trade and the World Trade Organization.

International strategic and financial developments in the late 1960s prompted the member states to move European integration in two new directions: monetary policy and foreign policy cooperation. In response to international currency fluctuations and Germany's overtures toward the Soviet bloc, national governments launched two separate initiatives. One aimed to achieve economic and monetary union by 1980; the other to coordinate member states' foreign policies as closely as possible. These were dramatic examples of deepening in the run-up to the EC's first enlargement, but European integration went through trying times in the years ahead. The reason was not enlargement itself, but the worldwide economic downturn that coincided with enlargement. In the recessionary 1970s and early 1980s, the economies of the member states diverged rather than converged; the ambitious and not well-thought-out goal of monetary union was scrapped; nontariff barriers proliferated in member states; and EC decision making ground to a halt.

The EC's fortunes revived in the mid-1980s when national governments, reeling from years of poor economic performance and facing intense interna-tional competition, rediscovered the benefits of market integration. One of the main architects of the EC's revival was none other than British prime minister Margaret Thatcher. Though later dismissed as a Euroskeptic, or zealous opponent of European integration, Thatcher championed the importance of completing the single European market by ensuring the free movement of goods, services, capital, and people. Thatcher looked back to the future: to the original provisions of the treaty establishing the EC (the Treaty of Rome) as a means of breaking down economic barriers among member states in order to boost growth and employment. Thatcher was instrumental in forging a consensus across the political spectrum in favor of what became known as the single market program. Optimistic assessments of the likely economic impact of a borderless EC, notably the Cecchini Report of 1988, bolstered her position.

Whereas Thatcher did not think that completing the single market necessitated institutional reform in the EC, most of the other national leaders wanted to enshrine their commitment to the single market in a new treaty, which would include provisions for greater use of qualified majority voting, an instrument of supranational decision making. This was the genesis of the Single European Act, the first major overhaul of the Treaty of Rome. Negotiation of the Single European Act coincided with another round of EC enlargement, Greece having joined in 1981 and Portugal and Spain being about to join in 1986. In keeping with the recurring fear that wider would mean weaker, especially as the new member states were substantially poorer than the existing members, national leaders included in the Single European Act a promise to strengthen regional (cohesion) policy by substantially increasing spending on development projects in the EC's disadvantaged areas. Thatcher saw this as a betrayal of the liberal principles of the single market program and a shameful side payment to the EC's poorer member states. By contrast, most other national leaders, as well as Jacques Delors, the commission's new president, saw it as a manifestation of solidarity in a growing EC and an acknowledgement that, contrary to neoliberal belief, market

forces alone would not guarantee the equitable or eventual spread of economic growth from the rich core to the poor periphery.

Thatcher and Delors fought their ideological battles in a series of speeches on the nature and purpose of European integration, delivered in the late 1980s. Embittered by the acceleration of integration beyond the single market program, and especially by the renewed momentum for economic and monetary union, Thatcher became increasingly isolated within the EC. Undisguised opposition to German unification set Thatcher further apart from her European counterparts. Battling fierce opposition at home to a proposed poll tax and growing hostility in Europe, Thatcher was forced from office by her own Conservative Party in 1990. Thatcher's departure and Delors's preeminence symbolized the remarkable transformation of the EC by the end of the 1980s, a period of profound change in the international system. With the end of the Cold War and the imminent collapse of the Soviet Union, a united Germany and a uniting Europe seemed to portend the triumph of European integration. As if to prove the point, the Maastricht Treaty of 1991, another far-reaching reform of the original Treaty of Rome, subsumed the EC into the EU, while pointing the way toward monetary union by the end of the 1990s, strengthening cohesion policy, launching a fledgling Common Foreign and Security Policy, and setting the stage for close cooperation in the policy areas of justice and home affairs.

The Overreaching EU It soon became apparent that the single market program was the zenith of European integration and that the Maastricht Treaty marked a profound turning point. The rejection of the treaty by a narrow majority of Danes in a referendum in June 1992 symbolized the EU's troubles. The result astounded Europe's political elites: how could a majority of citizens in any member state have rejected a carefully negotiated treaty intended to deepen European integration? Although there were many reasons why people voted against the treaty, the result reflected widespread public concern about the rapid pace of integration and, especially, the so-called democratic deficit the seeming unaccountability and remoteness of EU institutions. Aware of the danger of the democratic deficit, EU leaders had already addressed the issue by giving more legislative power to the directly elected European Parliament and promoting the principle of subsidiarity or states' rights. Most Europeans were unimpressed by the European Parliament, however, as evidenced by the declining turnout in successive elections. Nor has subsidiarity, which is difficult to put into practice, grabbed the popular imagination.

Public disaffection with the EU has increased steadily since the early 1990s. Subsequent treaty changes Amsterdam in 1997, Nice in 2001, and the Constitutional Treaty, signed by national leaders in 2005 failed to allay concerns about the democratic deficit or to endear the EU to the people. At the same time, poor economic performance in the big, continental member states further undermined public support, although responsibility for sluggish growth lay primarily with national leaders rather than the EU itself. Indeed, national leaders agreed at an EU summit in 2000 to undertake a series of labor market and other reforms the so-called Lisbon Strategy but governments in France, Germany, and Italy, in particular, were loath to antagonize volatile electorates by following through with concrete measures. The road to economic and monetary union in the 1990s may have made matters worse, as governments strove to reduce deficits and debts in order to meet the criteria for participation in the third stage of the process: adoption of a single monetary policy and a single currency. Regardless of its economic or political merits, the advent of monetary union in 1999 followed by the introduction of the euro in 2002 lacked popular enthusiasm. Most people in the eurozone miss their national currencies far more than they value the euro.

Europeans are especially concerned about enlargement, fearing that further expansion of the EU, following the accession of 10 Central and Eastern European states in 2005 7, is unsustainable. Economic arguments about the long-term benefits of enlargement larger market size, economies of scale, increased competition run counter to widespread

worries about worker migration from east to west, downward pressure on wages, and possible job losses. Under the circumstances, Turkey's application for EU membership seemed doomed by late 2007. The accession negotiations, which formally began in 2004, were overshadowed in any case by disputes between both sides over the future of northern Cyprus and over the pace and durability of political reforms in Turkey. Lurking in the background was the generally unspoken but politically potent viewpoint that Turkey, a predominantly Muslim country located mostly across the Bosporus, is neither culturally nor geographically European.

The fate of the Constitutional Treaty, a political initiative to improve the EU's functioning and fortunes, illustrates the state of the EU itself. Drafted by a convention representing a wide range of political opinion from the existing members and candidate countries, the Constitutional Treaty included provisions to improve decision making in an enlarged EU, address concerns about the democratic deficit, and enhance the EU's ability to act internationally. Yet sizable majorities in France and the Netherlands, two of the EU's founding member states, rejected the Constitutional Treaty in referendums in mid-2005. Their reasons for doing so, like those of the Danes who rejected the Maastricht Treaty in 1992, were varied, but included a strong dislike of the EU's political pretensions. After a lengthy "period of reflection," EU leaders redrafted the Constitutional Treaty dropping the word "constitution" or "constitutional" from the title and text and making other symbolic changes and signed the Lisbon Treaty in December 2007. Until ratified by all member states, the Lisbon Treaty cannot come into effect. In the meantime, the EU is operating under the terms of the existing treaties, whose institutional provisions are hardly adequate for such a large and multifaceted organization.

Prospects The EU has been in existence, in one form or another, since 1952. During that time it has grown from 6 to 27 member states and in scope from a common market for coal and steel to a single market in goods, services, capital, and people, as well as a monetary union for countries that meet the convergence criteria. Managing and perfecting the single market facilitating free movement within its borders is the core activity of the EU. It is an activity that requires constant vigilance, regulatory fine-tuning, and occasional intervention by the Court of Justice. It is also an aspect of the EU that enjoys public support, as Europeans have come to appreciate the benefits of unimpeded cross-border travel and shopping. Europeans appreciate as well the need for their governments to cooperate closely in the fight against terrorism and transnational crime, and to maximize their leverage by acting collectively on the global stage. Beyond that, Europeans appear to have little appetite for further enlargement and for extravagant constitutional schemes. There is a growing consensus that, at this stage in its history, the EU needs to focus on first principles and strengthen its fragile legitimacy by demonstrating its effectiveness in those policy areas in which it provides obvious added value.

See also common market; customs unions; European Central Bank; European Monetary Union

FURTHER READING

El Agraa, Ali M., ed. 2004 *The European Union: Economics and Policies*. 7th ed. Upper Saddle River, NJ: Prentice Hall. A good collection of essays by leading authorities on a range of economic theories, debates, and policies related to the European Union.

Artis, Michael, and Frederick Nixson. 2002. *The Economics of the European Union: Policy and Analysis*. 3d ed. Oxford: Oxford University Press. A good account of economic policies and policymaking in the EU.

Baldwin, Richard. 2006. *The Economics of European Integration*. 2d ed. New York: McGraw Hill. A solid, accessible text on economic integration in Europe.

Calingaert, Michael. 1988. *The 1992 Challenge from Europe*. Washington, DC: National Planning Association. A classic assessment of the implications of the single market program for the United States.

. 1996. *European Integration Revisited: Progress, Prospects, and U.S. Interests*. Boulder, CO: Westview. A follow up to Calingaert's 1988 assessment of the implications of the single market program for the United States.

Cecchini, Paolo. 1988. *The European Challenge, 1992: The Benefits of a Single Market*. Aldershot, U.K.: Gower. A synopsis of the multivolume report by Paolo Cecchini on the "cost of non Europe"   the cost of not integrating the European market.

Dinan, Desmond. 2004. *Europe Recast: A History of Euro pean Union*. Boulder, CO: Lynne Rienner. A compre hensive history of European integration.

———. 2005. *Ever Closer Union: An Introduction to Eu ropean Integration*. 3d ed. Boulder, CO: Lynne Rienner. A classic textbook on the European Union, now in its third edition.

Eichengreen, Barry. 2006. *The European Economy since 1945: Coordinated Capitalism and Beyond*. Princeton, NJ: Princeton University Press. An excellent history of economic policymaking in Europe from the end of World War II to the early 21st century.

Gillingham, John. 2003. *European Integration, 1950 2003: Superstate or New Market Economy?* Cambridge: Cam bridge University Press. A controversial book that characterizes the history of European integration as a struggle between economic centralizers and market re formers.

Molle, Willem. 2006. *The Economics of European Integra tion: Theory, Practice, Policy*. 5th ed. Aldershot, U.K.: Ashgate. One of the most popular and authoritative texts on the subject.

Neal, Larry. 2007. *The Economics of Europe and the European Union*. Cambridge: Cambridge University Press. A new version of a book with the same title originally written by Neal and Daniel Barbezat, published by Oxford Uni versity Press in 1998, that examines the EU economy collectively and the economies of the member states separately.

Pelkmans, Jacques. 2006. *European Integration: Methods and Economic Analysis*. 3d ed. London: Financial Times Press. A dense textbook on the economic principles and practices underlying European integration.

Wallace, Helen, William Wallace, and Mark Pollack. 2005. *Policy Making in the European Union*. 5th ed. Oxford: Oxford University Press. An authoritative and compre hensive examination of key EU policy areas and of the policymaking process.

DESMOND DINAN

■ evolution of development thinking

In the international community, a focus on devel opment emerged in the 1940s and 1950s, follow ing the focus on reconstruction and development after World War II and the subsequent decoloniza tion of former empires. As academics and senior civil servants turned their attention to how the newly decolonized regions could catch up to their richer counterparts, the arguments of development think ers soon became caught up in the schisms of the Cold War.

In the heat of the postwar period, the spec trum of solutions offered by development thinkers ranged from solutions inspired by Marxist thought to those that advocated closer integration in the capitalist system. In the 1950s and 1960s, however, even the procapitalist economic orthodoxy placed the state at the heart of economic development, and the dominant paradigm across the spectrum was of state-led growth. For development thinkers, the lesson of the theories of John Maynard Keynes and the chastening experience of the Great Depres sion was that self-regulated markets could not be relied on to generate prosperity. This perception had been reinforced by what were regarded as the relatively effective planning processes of post World War II Europe. Meanwhile, the small size and often enclave- or state-dependent private sec tor in developing countries meant that develop ment thinkers had little evidence of, and heard only muted arguments for, more effective and open markets.

The broad aim of development thinkers was to raise growth in poorer countries to achieve the same income levels, on average, as developed countries. At first, even the non-Marxist development thinkers looked for universal models of development and theories that would initiate the growth process. The idea that growth, once initiated, would continue through virtuous circles or reinforcing effects gained wide currency. This was echoed in development thinkers' phrases that persisted into the 1960s, such as the "big push" (Rosenstein-Rodan), the "takeoff" (Rostow), and the "great spurt" (Gerschenkron) (see Hirschman 1981).

Development practitioners and thinkers observed that market failures were particularly prevalent in developing countries. The private sector, which was relatively poorly developed and driven by narrow interests, had shown itself to be an inadequate engine of growth. In newly decolonized Africa and in Asia and Latin America, new national states were determined to establish postcolonial models of development. Given that the majority of professionals were concentrated in public service and had limited knowledge of the private sector, and that the absence of democracy meant that the state exercised virtually all the levers of power, the preoccupation with the role of the state was not surprising.

Two strands of thought emerged, both of which reflected this heavy reliance on the state. The first was the dominant strand within developing countries, particularly on the political left. This was import-substitution industrialization (a policy intended to promote industrialization by protecting domestic producers from competition from imports), which informed government thinking in Latin America, Africa, and Asia. The second strand of thought was multilateralism (a policy of multiple nations working cooperatively to solve world problems), which was given added impetus in the strategic agenda of Western Europe and North America.

Industrialization plans aimed at getting around the problems of coordination of industry development became the dominant currency of development thinkers residing in developing countries and some developed countries, particularly the United Kingdom. Planning models, at times elaborated through input-output matrixes, provided a vision of an integrated chain in which domestically produced raw materials and local labor provided the building blocks of industries that met first domestic and then export requirements. To the extent that the concept of dynamic comparative advantage influenced thinking, it was within the constraint of perceptions regarding the unequal terms of trade and protectionism in rich countries, which were seen as effectively preventing the balanced integration of developing countries into a global trading system. Greater self-reliance and import-substitution indus-

trialization, reinforced by fixed exchange rates, import tariffs, and exporting cartels, were seen by many as the appropriate response. These would prime the pump for a virtuous circle of increased employment and incomes, which would be coupled with improved health and education as well as rising consumer demand and investment. To the extent that the rich countries were seen to have a role, it was primarily as providers of capital and technology.

These ideas reflected the view that the prices for agricultural goods, expected to be primarily exported by developing countries, would fall in the long run and that the technological spillover effects from manufacturing were greater than those from agriculture (Prebisch 1950). Raúl Prebisch not only exercised a powerful influence in the evolution of development thinking but also was instrumental in the establishment of institutions that continue to carry his ideas forward, notably the Economic Commission for Latin America (ECLA) and the United Nations Conference on Trade and Development (UNCTAD).

The oil price increases of the 1970s also had important impacts on ideas about trade policy and development. Many developing countries considered that cartelization, as used by the Organization of the Petroleum Exporting Countries (OPEC), could be extended to other commodities to ensure that the gains from trade were more equally shared with developing countries. Government intervention at the national and supranational levels in markets and trade became a key element of what some termed the "New International Economic Order." UNCTAD and an array of marketing boards at the national and international levels (including the international coffee, cocoa, sugar, and commodity bodies established in the 1960s and early 1970s) were seen as providing the means to rebalance the international trading system for the benefit of developing economies.

While some thinkers continued to pursue global solutions, by the late 1970s many development specialists saw the evident constraints faced by bankrupt governments and narrowed their ambitions. Focus shifted toward filling needs such as those

for food, education, and health and away from overall integrated frameworks that attempted to incorporate all sectors of an economy. This evolution largely reflected a lack of success in kick-starting growth and a growing realization that theories did not account for political economy factors related to institutions, governance, and conflict. There was also evidence that the benefits of growth were not necessarily trickling down to the poorest people in developing countries and that some types of growth were environmentally unsustainable.

Development thinkers recognized that state-led models, at least in the medium term, had failed to address poverty and that stubborn levels of high unemployment were endemic in many societies. Chenery and others argued that policies should be devised to specifically benefit the poor (Chenery et al. 1979). The Basic Needs Approach was adopted by the International Labor Organization and later by the World Bank. Rather than challenging the role of governments, it sought to ensure that governments adopted pro-poor policies that would ensure that minimum needs would be met to enable all citizens to lead productive lives. In the 1980s, the capability approach developed by Amartya Sen and others built on this understanding with the United Nations Development Program in its 1990 Human Development Report. This report went beyond the analysis of the World Bank's annual World Development Reports to provide a broad understanding of the issues and data required to address poverty (Stewart 2006). The shift in development aims was accompanied by an expanded set of measures by which to track development. In particular, the human development indicators published in 1990 covered poverty, health, education, and the status of women in many of the world's countries. A number of these indicators, as well as others covering environmental sustainability, are now part of the Millennium Development Goals.

Primacy of Markets By the late 1970s, the oil crises, interest rate hikes, and widespread macroeconomic instability had led to a reconsideration of the balance of state and market. Growing awareness that the pressure on public-sector finances was not politically sustainable led to a reconsideration of governments' roles. The tension between the informational requirements of central planning and the increasing complexity and integration of developing economies was also becoming more evident. Macroeconomic instability associated, at least in part, with spikes in oil and interest rates led to deep crises, which in turn were seen by a new wave of development thinkers as symptomatic of a deeper malaise. By the 1980s, the global economic slowdown and the debt crisis had increasingly constrained the potential for state action in initiating development, and the voices of those who argued that governments should "get out of the way of development" were being heard more widely.

In the late 1970s, a new cohort of development thinkers, based mainly in the United States, was in the ascendant. They argued that state planners could not possibly possess all the knowledge required to make decisions reflecting efficiency as well as people's differing preferences. An influential World Bank study (Krueger et al. 1988) argued that import substitution often resulted in inefficient industries and an antiagriculture bias, undermining both growth and poverty reduction. Moreover, governments had revealed themselves to be collections of interests rather than the benevolent arbiters of the collective interest that the import-substitution industrialization model required. The development thinkers at the fore in the 1980s eschewed state-led development and focused on the removal of regulations, tariffs, and other government-induced economic and trade distortions. Governments were now viewed as a barrier to entrepreneurs and development, rather than as initiators of development. It was argued that the private sector, left to its own devices, would be the engine of economic growth.

The increasing market focus of development thinkers and practitioners reflected in part the economic experiences and the broader political climate of the time. Global economic growth slowed following the oil price shocks of 1974 and 1979. The second oil price shock was followed by higher U.S. interest rates to control inflation, which pushed the United States into recession. With falling demand

for their exports and rising interest rates, many developing countries could not service the debts they had built up over the preceding ten years, resulting in the debt crisis that began in Mexico in 1982 and spread to other developing countries. The belief that government officials and bureaucrats would act in their own interest led many thinkers to push for rapid change. Jeffery Sachs and others argued that a crisis would provide only a limited amount of time for reform. Reforms would have to be pushed through quickly, despite attendant costs, to ensure that the rent seekers who sought to gain wealth through benefiting from the imbalances and lack of transparency in the legal and regulatory environment could not reestablish their interests (Sachs 1994; Lal 1997).

In the richer members of the Organisation for Economic Co-operation and Development (OECD), the oil shock and subsequent rise in interest rates and debt servicing costs also led to a growing appreciation for the need for fiscal adjustment. With the dominant political mood more open to the arguments of those economists who called for reductions in pubic expenditure, policies became more promarket in key developed economies.

The broader geopolitical context also contributed to this pendulum swing. The hardening of the Cold War caused a reaction against state-led growth models, with these models being seen as part and parcel of socialism. In this context, advanced economies such as the United States, Britain, and Germany began the wave of deregulation and privatization under the leadership of Ronald Reagan (United States), Margaret Thatcher (United Kingdom), and Helmut Kohl (Germany). Given the power that these governments exercised over the Bretton Woods institutions, and that the trend reflected a broad shift in economic thinking, it is not surprising that in this period the strongly promarket Anne Krueger replaced the long-serving development economist Hollis Chenery in the influential role as chief economist of the World Bank.

In this environment, Williamson (1990) identified a set of policies that he termed the "Washington consensus." These policies reflected the prevailing view of orthodox development thinkers, not least at the Washington-based World Bank and International Monetary Fund, in the 1980s and the first half of the 1990s. These policies included macroeconomic stabilization; liberalization of trade, exchange rates, and financial markets; deregulation of domestic industries; and privatization of government-owned businesses. What Williamson identified was not only the prevailing Washington view. It also was applied by a growing number of developing countries, often under the duress of fiscal adjustment, as these countries sought a prerequisite to sustainable growth.

The consensus broke down over the extent of liberalization required for growth. For the Washington-based multilateral lending institutions, this required the removal of regulations including trade barriers, freeing prices from controls, and the dismantling of marketing boards and similar state institutions to allow markets to work to achieve their reputed efficiency. Like the classical economists of the 19th century, these thinkers advocated a highly restricted role for the government involving the provision of public goods such as law and order, defense, and a sound currency, and when necessary, primary education, primary health care, and a social safety net.

The late 1980s saw improvements in macroeconomic stability and an opening to trade, financial, and other flows in much of the developing world. From the 1980s to 2005, trade barriers fell across developing economies from average levels of more than 30 percent to about 10 percent. The "tiger economies" of East Asia were held up as examples of the benefits of getting rid of government regulation and controls, promoting exports, and "letting markets work," and several studies appeared to confirm that these policies were responsible for the rapid growth of these economies (World Bank 1993). These studies were later shown to provide an inadequate explanation of the success of the "Asian Tigers." The interpretation of the industrialization of Asian economies has been challenged by a number of critics who argued that the state played a pivotal role in development (Amsden 2001; Wade 1990), and even the World Bank later developed a more nuanced

view of the East Asian experience (Stiglitz and Yusuf 2001).

Importance of Institutions By the late 1990s, the failure of the rapid transition from centrally planned to capitalist economies to deliver broad-based benefits in Eastern Europe, together with the Asian crisis of 1997, in which institutional weakness played a part, and a growing perception that the distribution of globalization's gains was inequitable, led to a widening acknowledgment of the shortcomings of a purely "promarket" approach to development. In particular, it neglected the importance of institutions and the provision of public goods, not least good governance, both at the national and at the global level. The pendulum shift in development thinking built on the work of the 1993 Nobel laureate Douglass North, who noted that individual behavior was constrained by numerous rules and social norms of the society in which the individual lives (North 1990). It was argued that without studying these constraints, including how they arise and their impacts, attempts to promote development would be unlikely to succeed.

The constraints provided by history became particularly obvious following the attempted shock therapy in the former Soviet states. In contrast, China continued to implement market-oriented reforms while nevertheless holding to its core commitment to a powerful state and with only partial reform of key macroeconomic and institutional levers.

Development practitioners began promoting stable and effective government institutions, strong enforcement of property rights, the absence of bureaucratic harassment, a lack of corruption, and protection from organized crime all things that matter for the functioning of markets and that had been previously neglected. In retrospect it was now considered naïve to think that the institutions that enable markets to work could evolve quickly and without state intervention.

The Asian crisis, Soviet transition, and Chinese experience highlighted the importance of sequencing, the order in which policies were put into effect. The orthodox position promoting liberalization of all aspects of the capital account was replaced by a heterodox view that highlighted the importance of sequencing and country specificity.

Complementarity of Markets and Governments By the mid-1990s, the dominant development discourse had begun to move away from market-led approaches. In its place, a consensus was forming around the complementarities between markets and governments. Attention focused on how government and the private sector may best work together. A vibrant private sector was now seen as the driver of economic growth, but this required properly functioning state institutions to build a good investment climate and deliver basic services competently. Investment climate analysis was focused on the ease and cost of starting legitimate businesses, including the bureaucratic costs and access to credit and basic infrastructure and utilities. Indicators based on World Bank surveys provide a comparison of the performance of different countries (World Bank Doing Business Project and World Bank Investment Climate Surveys).

Countries that combined institutional improvements with market-oriented policy reforms and greater engagement with the world economy saw their per capita incomes grow in the 1990s at the historically very rapid pace of more than 5 percent per year. Some countries have achieved even faster growth, with China's per capita growth averaging 8.7 percent from 1990 to 2005. That a sound state is required for solid growth was shown in the uneven development outcomes of the former Soviet countries. Promarket policies could not be expected to succeed without strong market-based institutional foundations.

Important too is local "ownership" of the development agenda rather than imposition from outside. The countries that have achieved rapid development, such as China and India, have done so through the actions of their own people and government, and with very little development assistance.

Growth Theory and Development Thinking Growth has been a concern throughout the history of development thinking, albeit with a general decline in expectations of what development assistance and policies can do. Yet growth theory and devel-

opment thinking diverged considerably in the 1960s, with growth theory identifying formal mathematical models these were mainly focused on growth in advanced economies. Underlying these models were assumptions of well-functioning markets and a system of economic incentives.

In contrast, development theorists typically eschewed mathematical models and focused on the failure of the assumptions used in growth theory. Interest in input-output and later general equilibrium analysis has provided a heavily quantitative seam within the development literature, but such methods remain a relatively minor feature of development thinking (Goldin, Knudsen, and Brandao 1994).

With increasing macroeconomic stability has come a greater recognition of the importance of microeconomic and institutional development. To the extent that the focus of many development thinkers today is mainly on how to affect the actions of individuals and firms rather than on macroeconomic aggregates, interest in models based on broad aggregates of national accounts has diminished.

Both growth theory and development theory have had a considerable influence on development policy. The simplistic stories that emerge from growth models have encouraged theories of missing components of production and policies that fill these gaps. These have variously included capital, foreign exchange, human capital, and technology. Development thinkers have been more active than growth theorists in exploring the reasons why these factors may be missing. In this sense, recent development thinkers have focused on the incentives for people to accumulate physical and human capital, and on an efficient use of these types of capital.

Questions regarding the sustainability of growth have only recently begun to enter the mainstream of development theory. The oil price shocks of the 1970s and the rapid recovery of the 1980s, not least in East Asia, led to a growing recognition of the need to focus on resource degradation and what was termed (by the Club of Rome) as the Limits to Growth. The United Nations Commission on Environment and Development in 1987 published the widely cited Brundtland report, which defined sustainable development as meeting the needs of the present without compromising the ability of future generations to meet their own needs (WCED 1987). It would be 20 years, however, until the evidence of climate change became so overwhelming that the largely grassroots movement influenced by the Brundtland report could see sustainability becoming central to development thinking.

No Single Model for Development Development thinking has evolved rapidly since the 1950s and is continuing to change. Different development concepts have emerged in response to specific economic and political conditions and problems, attempting to address these challenges, but typically only partially succeeding. Concepts have been less useful when they have been exported without accounting for different economic, social, and cultural conditions, or when they have continued to be applied after economic and political conditions have changed. Development strategies have been most successful when they addressed the specific challenges facing a country at a particular point in time. This requires that countries have the capacity for policy formation in terms of skilled researchers and the relevant data. Both are vital to policy reform. Since the mid-1990s, development thinking has continued to evolve. Increasingly, development is seen not through the lens of actual outcomes but through the outcomes that a person is able to choose. This concept gives greater prominence to empowerment of the individual and the expansion of individual freedom; and individuals are seen as agents for change rather than as passive recipients of assistance.

It has become clearer since the 1950s that there is no development panacea or universal path for development. Development thinkers and practitioners can benefit from understanding the experiences of different countries and by drawing on the latest theory. There is, however, no shortcut. If anything, the evolution of development thinking shows how today's truths are replaced by tomorrow's wisdom. The application of development thinking requires an arsenal of knowledge that includes economic theory and is informed by comparative and

historical analysis. The application also requires that the thinking respond to a country's specific priorities and circumstances. Greater investment in research is required to ensure that development thinking evolves even more rapidly in response to the enormous current challenges, many of which, such as global equity and climate change, go the heart of questions of sustainable development.

See also aid, international; development; economic development; Millennium Development Goals; poverty, global; Washington consensus

FURTHER READING

Adelman, I. 2001. "Fallacies in Development Theory and Their Implications for Policy." In *Frontiers of Development Economics: The Future in Perspective*, edited by G. M. Meier and J. E. Stiglitz. Oxford: Oxford University Press, 103 34. Discusses the impact of the narrow preoccupation of economics in the field of development economics.

Amsden, A. 2001. *The Rise of "The Rest" Challenges to the West from Late industrializing Economies.* New York: Oxford University Press. Challenges convergence theory and examines models of late industrialization, comparing the "independent" model of China, Korea, India, and Japan, and the "integrationist" model of Argentina, Brazil, Mexico, and Turkey.

Bruton, H. J. 1998. "A Reconsideration of Import Substitution." *Journal of Economic Literature* 36 (2): 903 36. Provides a critical view of import substitution and contrasts it to the "outward oriented" approach to development.

Chenery, H., et al. 1979. *Redistribution with Growth.* Oxford: Oxford University Press. Coauthored with leading development thinkers of the time, this is a key statement of World Bank development strategy under Hollis Chenery.

De Soto, H. 2001. *The Mystery of Capital.* London: Black Swan. Argues that property rights and other reforms are required for the poor to turn "dead" into "liquid" capital.

Goldin, I., O. Knudsen, and A. Brandao. 1994. *Modeling Economywide Reforms.* Paris: OECD. Examines the impact of macroeconomic and sectoral reforms using a variety of economic models, comparing methodologies and outcomes.

Goldin, I., and K. Reinert. 2006. *Globalization for Development: Trade, Finance, Aid, Migration, and Policy.* Washington, DC: World Bank and Palgrave Macmillan. Examines the key economic flows and shows how policies at the national and global level are required to make globalization work for development.

Hirschman, A. O. 1981. *Essays in Trespassing: Economics to Politics and Beyond.* Cambridge: Cambridge University Press. A collection of essays providing an overview of the author's scholarship, including work on economic development and its political repercussions, as well as on political participation, rational choice theory, and the history of economic thought.

Krueger, A., M. Schiff, and A. Valdes. 1988. "Agricultural Incentives in Developing Countries." *World Bank Economic Review* 2 (3): 255 71. Estimates the impact of sector specific (direct) and economywide (indirect) policies on agricultural incentives for 18 developing countries for the period 1975 84.

Lal, D. 1997. *The Poverty of "Development Economics."* London: Institute of Economic Affairs. Argues for market based economic management and less government intervention to address development challenges.

Lindauer, D. L, and Lant Pritchett. 2002. "What's the Big Idea? The Third Generation of Policies for Economic Growth." *Economia: Journal of the Latin American and Caribbean Economic Association* 3 (1):1 28. Puts the Latin American experience of reform in global context and offers policy advice on economic growth.

Myrdal, G. 1957. *Economic Theory and Underdeveloped Regions.* London: Gerald Duckworth. Based on a series of lectures delivered in Cairo in 1955, examining the basic principles of classical economic theory (particularly international trade) in relation to the problems of underdeveloped regions.

New Palgrave: A Dictionary of Economics. 1989. "Economic Development." Basingstoke: Palgrave Macmillan. Compendium of economics concepts and leading theorists.

North, D. 1990. *Institutions, Institutional Change, and Economic Performance.* Cambridge: Cambridge University Press. Develops an analytical framework for explaining the ways in which institutions and institutional change affect the performance of economies.

Prebisch, R. 1950. *The Economic Development of Latin America and its Principal Problems.* New York: UN Economic Commission for Latin America. Challenges comparative advantage as a basis for development, ar guing that developing countries (the "periphery") are in a structurally unequal relationship with industrialized countries (the "center"), and that nonmarket solutions at the national and global level as a consequence are required to address development challenges.

Ray, D. 1998. *Development Economics.* New Delhi: Oxford University Press. Synthesizes literature in the field and raises policy questions, arguing that there is no single cause for economic progress, but that a combination of factors (physical and human capital, reduced in equality, and strong institutions) consistently favor de velopment.

Rostow, W. W. 1990. *Theorists of Economic Growth from David Hume to the Present.* New York: Oxford Uni versity Press. Explores how growth theorists since the mid 18th century have dealt with variables and prob lems posed by the dynamics of economic growth.

Sachs, J. 1994. *Understanding Shock Therapy.* New York: Social Market Foundation. Motivation for rapid ad justment of the former Soviet Union to a market econ omy, and the need for radical reform of prices, liber alization, and privatization.

Sen, A. 1999. *Development as Freedom.* New York: Knopf. Argues that open dialogue, civil freedoms, and political liberties are vital dimensions of sustainable development.

Stern, N. H., J. Dethier, and F. Halsey Rogers. 2005. *Growth and Empowerment: Making Development Hap pen.* Cambridge, MA: MIT. Analysis of the interaction between growth and poverty reduction which empha sizes the importance of focusing on the "twin pillars" of investment and empowerment.

Stewart, F. 2006. "Basic Needs Approach." In *The Elgar Companion to Development Studies*, edited by D. Clark. Cheltenham: Edward Elgar, 14 18. Provides a synthesis of the basic needs approach and places it in the context of the evolution of development thinking.

Stiglitz, J. 2000. "Capital Market Liberalization, Economic Growth, and Instability." *World Development* 28 (6): 1075 86. Drawing lessons from the Asian crisis, high lights the importance of sequencing and of a nuanced approach to capital market liberalization.

Stiglitz, J., and S. Yusuf, eds. 2001. *Rethinking the East Asian Miracle.* New York: Oxford University Press. Compares the industrial policies of East Asian countries, including China, and sheds new light on the "East Asian Miracle."

Wade, R. 1990. *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization.* Princeton, NJ: Princeton University Press. Examination of industrial policy in East Asia focusing on the inter action of state and market.

Williamson, J. 1990. "What Washington Means by Policy Reform." In *Latin American Adjustment: How Much Has Happened?* Edited by J. Williamson. Washington, DC: Institute for International Economics, 5 20. The text that gave rise to the notion of the "Washington con sensus" and which summarizes the orthodoxy of the time.

Wolf, M. 2003. "The Morality of the Market." *Foreign Policy* 138: 46 50. Argues for the ethical and equity basis of capitalism on the basis of superior outcomes.

World Bank. 1993. *The East Asian Miracle: Economic Growth and Public Policy.* New York: Oxford University Press. World Bank policy report which argues that strong macroeconomic policies, a focus on the market, narrowly targeted government interventions, and phys ical and human capital account for the success of the East Asian economies.

World Commission on Environment and Development (WCED). 1987. *Our Common Future.* Oxford: Oxford University Press. United Nations report that provided the first global attempt to locate environmental prob lems in the context of growth and development and to identify the long term implications of rising consump tion and resource depletion.

IAN GOLDIN

■ exchange controls
See capital controls

■ exchange market pressure
Foreign exchange market pressure (EMP) indexes are weighted schemes of variables designed to gauge

speculative pressure on a country's currency. These variables normally include the exchange rate, the domestic interest rate, and/or a measure of international reserves available to the monetary authority. Crisis incidents in recent years, including the severance of the British pound from the exchange rate mechanism in 1992, the Asian crisis in 1997, and the abolition of the Argentinean peg with the U.S. dollar in early 2002, have shown that the potential for foreign exchange market pressure to result in currency crises is significant, and hence monitoring informative variables is crucial.

Investors may put downward pressure on a currency for several reasons: for example, they may have spotted a fundamental imbalance and believe that a given exchange rate is unsustainable; or they may observe and follow the actions of other investors/ speculators. The reaction of policymakers depends on whether they want to accommodate the pressure and allow devaluation or, assuming a tough stance, they prefer to deter it. If the latter is the case, they may increase the short-term interest rate (hence offering a greater return than before on domestic currency investments) or spend international reserves (buying the domestic currency in the foreign exchange market with the aim of supporting its value). An EMP index is designed to capture this kind of activity in the foreign exchange market.

Theoretical Foundations The roots of the EMP index can be found in an analysis of demand and supply of national currencies in Girton and Roper (1977), where the term *exchange market pressure* was first used. A monetary model of the balance of payments is derived in which changes in the exchange rate and real international reserves for a small open economy are a function of changes in domestic credit expansion and real income, the money base and real income of a center currency country, as well as the inflation differential and the uncovered interest differential.

The widespread effects of the crises in the 1980s and the early 1990s highlighted the need to explain and predict such incidents. Girton and Roper's model provided the necessary insight for finding a solution to the problem of how to measure pressure

in the foreign exchange market. Whereas changes in the exchange rate can be easily used to capture the effects of a successful speculative attack, they cannot identify instances where in spite of speculative pressure being present no significant loss in the value of the currency has taken place. Furthermore, devaluation episodes may be voluntary on the part of policymakers attempting, for example, to boost exports. Clearly, in order to construct meaningful models of crisis prediction, also known as early warning systems, a robust definition of what constitutes a crisis needs to include a measure of international reserves.

Practical Issues The issue arises of how to weight the two variables exchange rates and reserves in an index of speculative pressure. One method commonly used is to employ the ratio of the standard deviation (σ) of the two components (Kaminsky and Reinhart 1999):

$$Index = \frac{\Delta e}{e} - \frac{\sigma_e}{\sigma_r}\frac{\Delta r}{r},$$

where $\Delta e/e$ stands for the percentage change in the nominal exchange rate and $\Delta r/r$ stands for the percentage change in the level of international reserves. The negative sign in front of the second term implies that losses in reserves are associated with a higher index value. Another weighting scheme could involve the inverse of the standard deviations. In addition, a third variable, namely, interest rate changes ($\Delta e \neq e$), is sometimes included in the computation. Thus another commonly used EMP index (Eichengreen, Rose, and Wyplosz 1995) is as follows:

$$Index = \alpha\frac{\Delta e}{e} + \beta\frac{\Delta i}{i} - \gamma\frac{\Delta r}{r},$$

where α is the inverse of the standard deviation of the percentage change in exchange rate divided by the sum of the inverses of the standard deviations of all three variables, β is the inverse of the standard deviation of the percentage change in the interest rate divided by the sum of the inverses of the standard deviations of all three variables, and γ is the inverse of the standard deviation of the percentage change in real reserves again divided by the sum of the inverses of the standard deviations of all three variables. This weighting scheme ensures that the most volatile

components do not dominate the index. The variables are expressed relative to a reference country.

Advantages and Problems The advantage of EMP indexes over simple measures of exchange rates is that even if the exchange rate does not move in the presence of speculative pressure, the pressure is registered through the policy response of the authorities, which might be an increase in interest rates and/or a reduction in reserve levels. Very often the indexes are converted into a discrete binary variable so that observations can be classified as either crisis or noncrisis. It is commonplace to allocate a one whenever a value of the index exceeds the mean of the index by three standard deviations. Otherwise, a zero is allocated. It needs to be noted that the definition of a critical threshold for classifying an observation as crisis is rather arbitrary; as a result, a wide range of critical thresholds has been used in the literature.

The identification of crisis episodes is also sensitive to other factors, for instance, sample selection and data frequency. Extending a given sample by, say, incorporating a period of excess volatility will push the critical threshold up (as a result of a higher standard deviation in the sample), which may lead to different results: observations that were characterized as crisis in the original sample may now turn into noncrisis observations. The frequency of the dataset used in the construction of the index is also of importance, as crises in the foreign exchange market can be short lived, with exchange rates, interest rates, and reserves assuming their precrisis values soon after an episode. It may be the case that if an annual frequency is selected, crisis incidents may be missed. Higher frequency data increase the likelihood of spotting a crisis but these are not always readily available for all the components of the index, especially data on interest rates. A further issue relates to the fact that different exchange rate regimes may be in place over the time span of the index (de jure or de facto). This can have implications for the index. For example, the switch from a fixed regime to a flexible one may be followed by a sustained depreciation of the domestic currency, with the index failing to readily convey this information (due to the lower variability in the interest rate and reserves changes).

Overall, despite their deficiencies, EMP indexes have been widely used in the context of early warning systems to inform policymaking. The absence of components relating to capital controls and exchange rate regimes implies that they should be used in conjunction with relevant information about the implementation of the prevailing exchange rate policies.

See also balance of payments; capital controls; contagion; currency crisis; early warning systems; European Monetary Union; exchange rate regimes; exchange rate volatility; foreign exchange intervention; interest parity conditions; international reserves; monetary conditions index; money supply; speculation; spillovers; sterilization

FURTHER READING

Eichengreen, Barry, Andrew Rose, and Charles Wyplosz. 1995. "Exchange Market Mayhem: The Antecedents and Aftermath of Speculative Attacks." *Economic Policy* 21: 251 98. Interest rates are incorporated for the first time in the definition of the foreign EMP index.

Girton, Lance, and Don Roper. 1977. "A Monetary Ap proach of Exchange Market Pressure Applied to the Postwar Canadian Experience." *American Economic Review* 67(4): 537 48. Laid the theoretical foundations for the construction of EMP indexes.

Kaminsky, Graciela L., and Carmen M. Reinhart. 1999. "The Twin Crises: The Causes of Banking and Balance of Payments Problems." *American Economic Review* 89 (3): 473 500. Uses a measure of EMP in order to assess an early warning system.

ALEX MANDILARAS

■ exchange-rate-based stabilization
See discipline

■ exchange rate forecasting
While exchange rate changes can lead to significant gains or losses to private entities involved in international transactions, from a macroeconomic perspective they can also have important impacts

on inflation and output growth. Exchange rate determination has therefore been a topic much researched by central banks and financial markets. Two broad methods of exchange rate determination fundamental analysis and nonstructural analysis have been used by economists. Based on the evidence, however, it appears that no single model can forecast exchange rates for all currencies or at all times.

Fundamental Analysis Since the mid-1980s, a multitude of fundamentals-based models have been used to assess the predictability of exchange rates. The host of fundamentals-based or structural exchange rate models includes monetary models, portfolio balance models, behavioral (fundamental) equilibrium exchange rate models, productivity differential models, the interest rate parity equation, and the purchasing power parity specification. One general observation is that there is no consistent evidence of a complex structural model outperforming a simpler one.

For economists, the monetary model, of which money market equilibrium is an important component, has been the workhorse model of exchange rate determination. In their seminal work, however, economists Richard Meese and Kenneth Rogoff (1983) forcefully show that forecasts from the monetary model do not outperform the naïve forecast from a random model that uses current exchange rates to predict future exchange rates. The result has largely withstood the challenge of sophisticated exchange rate models and elaborate econometric techniques developed after its publication. Consequently, the Meese and Rogoff study has greatly shaped the subsequent empirical studies on forecasting exchange rates.

More recent empirical work rejuvenates the monetary model by showing that its forecasting performance can beat the random walk's over long horizons (Mark 1995). The intuition is that, despite its large short-term volatility, an exchange rate's long-run value is determined by economic fundamentals (also see Chinn and Meese 1995). These long-horizon results are, nonetheless, not that persuasive once they are subjected to close scrutiny. Besides the choice of sample periods and the method of con-

structing fundamental values, an issue is the interaction between short- and long-horizon forecasts. If long-horizon forecasts are constructed from a sequence of short-horizon forecasts, then the performance of these two types of forecasts should be linked (Berkowitz and Giorgianni 2001). Further, forecasting into the distant future usually involves a high level of uncertainty. These considerations plus some more recent studies cast doubts on the results that long-horizon forecasts from structural exchange rate models are better than those from a random walk.

Despite some setbacks, long-horizon forecasts are still the focus of some foreign exchange forecast exercises. Efforts to revive the monetary model's forecast ability include the uses of long history data, panel data, nonlinear dynamics, and nonparametric techniques. Notwithstanding all of these efforts, an extensive study shows that, despite some evidence of forecast superiority, it is difficult to find a fundamentals-based exchange rate model outperforming a random walk specification for all the sample periods, evaluation criteria, and currencies under consideration (Cheung, Chinn, and Pascual 2005).

Nonstructural Analysis Given the dismal performance of structural exchange rate models, researchers have explored the forecasting ability of nonstructural time series models. Various time series techniques have been used to model exchange rate behavior. Some common ones are standard autoregression-moving-average models, vector autoregression models, vector error correction models, fractional integration models, threshold autoregression models, Markov switching models, and nonparametric models. Again, there are reservations about the positive results of using a time series model to forecast exchange rates.

Technical analysis that relies on price patterns is also commonly used to forecast exchange rates. Technical analysis uses a range of tools from the relatively simplistic moving average rules to complicated neural network based techniques. Despite the skeptics in academia, technical analysis is quite popular among professional foreign exchange traders (Cheung and Wong 2000). The performance of technical analysis has been quite unstable over time,

however, and there has been a debate on whether the observed profits from technical analysis are compensations for trading risk or excess (abnormal) returns (Menkhoff and Taylor 2006).

Summing Up Since Meese and Rogoff reported the failure to beat the forecast performance of a random walk, there have been serious efforts to determine the predictability of exchange rates. Sophisticated exchange rate models both structural and nonstructural and refined econometric techniques have been developed to challenge the random walk forecast. In the process of finding the real contender, various reasons have been offered to explain the difficulty in forecasting exchange rates. The most commonly cited reasons include the time-varying relationship between exchange rates and their fundamentals, bad data quality, and incorrect model specification.

With the intense investigation since the mid-1980s, we are still waiting for convincing evidence that an exchange rate model can consistently outperform the random walk. In sum, exchange rate forecasting is still an elusive exercise, and the extant evidence confirms the adage that no single model can forecast exchange rates for all currencies or at all times.

See also balance of payments; equilibrium exchange rate; exchange rate volatility; interest parity conditions; purchasing power parity; real exchange rate

FURTHER READING

Berkowitz, J., and L. Giorgianni. 2001. "Long Horizon Exchange Rate Predictability?" *Review of Economics and Statistics* 83: 81 91. Raised reservations about the reported positive performance of long horizon forecasts.

Cheung, Y. W., M. D. Chinn, and A. Garcia Pascual. 2005. "Empirical Exchange Rate Models of the Nineties: Are Any Fit to Survive?" *Journal of International Money and Finance* 24: 1150 75. Reported the difficulty of finding a model outperforming a random walk across currencies, forecasting periods, forecasting horizons, and evaluation criteria.

Cheung, Y. W., and C. Y. P. Wong. 2000. "A Survey of Market Practitioners' Views on Exchange Rate Dynamics." *Journal of International Economics* 51: 401 19. One of a few survey papers that documents, among other things, the popularity of technical analysis among foreign exchange traders.

Chinn, M. D., and R. A. Meese. 1995. "Banking on Currency Forecasts: How Predictable Is Change in Money?" *Journal of International Economics* 38: 161 78. Showed that long horizon forecasts from structural models can beat a random walk.

Engel, C., and J. Hamilton. 1990. "Long Swings in the Exchange Rate: Are They in the Data and Do Markets Know It?" *American Economic Review* 80: 689 713. Its result of a Markov model beating a random walk was called into question a few years later.

Mark, N. 1995. "Exchange Rates and Fundamentals: Evidence on Long Horizon Predictability." *American Economic Review* 85: 201 18. A noted exercise that, among other things, popularizes the examination of long horizon forecasts.

Meese, R., and K. Rogoff. 1983. "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics* 14: 3 24. Reported the difficulty in beating the forecast performance of a random walk.

Menkhoff, L., and L. P. Taylor. 2006. "The Obstinate Passion of Foreign Exchange Professionals: Technical Analysis." University of Warwick Working Paper. Warwick, UK. Surveyed the issues related to the use of technical analysis in foreign exchange markets.

YIN WONG CHEUNG

■ exchange rate pass-through

Exchange rate pass-through can be defined as the degree of sensitivity of import prices to a one percent change in exchange rates in the importing nation's currency. A closely related term is *pricing to market*, which refers to the pricing behavior of firms exporting their products to a destination market following an exchange rate change. More to the point, *pricing to market* is defined as the percentage change in prices in the exporter's currency due to a one percent change in the exchange rate. Thus the greater the degree of pricing to market, the lower the extent of exchange rate pass-through.

At one extreme, if import prices change by the same proportion as the change in the exchange rate, the result is full or complete pass-through and hence no pricing to market. At the other extreme, if exporters adjust prices in their own currency by the same proportion as the exchange rate change but in the opposite direction, the result is full pricing to market but no or zero pass-through of the exchange rate change to the destination market prices. More generally, if exporters alter the export prices in their own currency by a proportion smaller than the exchange rate change, then exchange rate pass-through is said to be partial or incomplete.

The degree of exchange rate pass-through and pricing-to-market behavior has important bearings on economic policy. If pricing to market is high and exchange rate pass-through low, then any exchange-rate-based adjustments to improve the trade balance for economies may be ineffective, as nominal exchange rate changes do not translate into real exchange rate changes.

Although these concepts have been well known to economists for a long time, they attracted particular interest following the Plaza Accord in 1985 and the subsequent sharp appreciation of the Japanese yen in relation to the U.S. dollar. Following this strengthening of the yen, other things being equal, one would have expected the unit price of Japanese products sold in the United States (in U.S. dollars) to have risen sharply. In reality, however, the price of Japanese cars and electronic items sold in the United States rose only marginally or remained constant, and in some cases actually declined (Goldberg and Knetter 1997). This suggested that the Japanese firms exporting products to the United States may have been absorbing a large part of the exchange rate changes in order to maintain market share. Given this important empirical observation, economists began trying to estimate the extent of exchange rate pass-through, as well as to analyze the determinants of exchange rate pass-through and the corresponding pricing-to-market behavior.

What Determines Exchange Rate Pass-Through? Among the most important factors that determine the extent of exchange rate pass-through are the size of the export market and the degree of competition the exporter faces in that market. If the export market for the product is large, then exporting firms are often willing to absorb a proportion of the exchange rate change so as not to lose market share. This is particularly true if the industry is highly competitive. The presence of a large number of suppliers selling similar goods in the market provides domestic consumers with a choice of many substitutes, making them relatively price sensitive. Conversely, if the industry is highly differentiated and exporters do not face much competition for their products, then exporter prices may be somewhat less responsive to exchange rate changes. In this situation, pricing to market will be lower and the corresponding pass-through will be higher. For example, exports to certain competitive industries in the United States, such as autos and alcoholic beverages, show relatively high pricing to market and corresponding lower exchange rate pass-through as exporters try to preserve market share (Knetter 1993).

The direction, duration, and magnitude of exchange rate changes also affect pass-through. If the currency of the destination market depreciates, then exporters may be willing to absorb this exchange rate change in order to keep local currency prices of their products stable and retain market share. In this situation, exchange rate pass-through may be low or incomplete. If the currency of the destination market strengthens, however, the exporter's product will be relatively cheaper and the exporting firm may engage in complete exchange rate pass-through. In other words, the response of exporters to exchange rate changes may be asymmetric, depending on whether the currency rate appreciates or depreciates.

The high costs of changing prices, as well as the possibility that frequent changes in unit sales prices (in the destination market's currency) can adversely affect a firm's reputation, may prevent firms from passing through temporary fluctuations in exchange rates. When exchange rate changes are large or appear to be permanent, however, exporting firms are more likely to pass through the changes to avoid a sharp reduction in their profit margins.

Low and Declining Exchange Rate Pass-Through Exchange rate pass-through generally has a greater effect on import prices than on a nation's consumer price index. This is because the latter includes nontradables that are less responsive to exchange rate changes. Regardless of the price index used, however, exchange rate pass-through was lower in the 1990s than in the 1980s, and has continued to decline. Although most of the research has focused on developed countries, where more data are available, some studies suggest that the conclusion holds for developing countries as well (Ghosh and Rajan 2006).

Exchange rate pass-through may also depend on a country's monetary and exchange rate policies. The more stable a country's monetary policy and the lower its rate of inflation, the lower the extent of exchange rate pass-through will be, as it is less likely that foreign exporters will pass through exchange rate changes (Taylor 2000). This in turn helps to sustain low inflation and makes monetary policy more effective. Thus there may be a "virtuous cycle" between stable monetary policy and low exchange rate pass-through.

If exports are invoiced in the currency of the importing nation known as consumer-currency pricing or local-currency pricing then exchange rate changes have little effect on the destination market import prices, which leads to low exchange rate pass-through. On the other hand, if exports are invoiced in the currency of the exporters referred to as producer-currency pricing then exchange rate changes have a greater effect on prices in the importing nation, leading to higher pass-through. It has been argued that if exporters set their prices in the currency of the country that has the more stable monetary policy (i.e., local-currency pricing as opposed to producer-currency pricing), then exchange rate pass-through into import prices in local currency terms will be correspondingly low (Devereux and Engel 2001).

In an important paper, the economists Campa and Goldberg (2005) test the significance of changes in macroeconomic variables and the extent of exchange rate pass-through into aggregate import prices for 25 industrial nations for the period 1975 99. The authors find that the lower the average rate of inflation and the less variable the exchange rate, the lower the corresponding extent of exchange rate pass-through will be. These macroeconomic factors play a minor role in explaining the low exchange rate pass-through, however, compared to the changing composition of a nation's imports away from raw materials and energy imports toward manufactured imports. This changing composition of the basket of imports toward manufacturing goods (which tend to be characterized as being more competitive industries) may be behind the low and declining rates of exchange rate pass-through.

Another factor that may have affected the extent of pass-through is cross-border "production fragmentation," which refers to the dispersion of the production process among different countries. A country might export the final product but at the same time import the corresponding parts and components from another nation. A depreciation of the currency of the nation exporting the final good makes its imported components more expensive. Therefore, an exchange rate change affects the exporter's costs, which leads the exporting firm to raise its price denominated in its own currency, and subsequently pass through less of the exchange rate changes into the partner nation's currency denominated prices. Moreover, with dispersion of production processes, often more than one nation supplies parts and components, thereby increasing competition and lowering exchange rate pass-through (Ghosh and Rajan 2006).

Policy Relevance The low exchange rate pass-through in the United States may explain the persistence of the U.S. trade deficit despite declines in the U.S. dollar. Conversely, low exchange rate pass-through implies that economies may be less concerned about the potential inflationary consequences of exchange rate fluctuations. Although the extent of exchange rate pass-through has important macroeconomic implications, it is predominantly a microeconomic phenomenon and depends significantly on the types of goods being traded.

See also expenditure changing and expenditure switching; fragmentation; New Open Economy Macroeconomics; nontraded goods; Plaza Accord; real exchange rate

FURTHER READING

Campa, J. M., and L. S. Goldberg. 2005. "Exchange Rate Pass Through into Import Prices." *Review of Economics and Statistics* 87 (4): 679 60. This widely cited paper tries to empirically determine the reasons behind the low and declining exchange rate pass through.

Devereux, M., and C. Engel. 2001. "Endogenous Currency of Price Setting in a Dynamic Open Economy Model." NBER Working Paper No. 8559. Cambridge, MA: National Bureau of Economic Research. Discusses how currency invoicing affects exchange rate pass through.

Ghosh, A., and R. S. Rajan. 2006. "A Survey of Exchange Rate Pass Through in Asia" *Asian Pacific Economic Lit erature* 21 (2): 13 28. An up to date review of empirical studies on exchange rate pass through in various Asian economies.

Goldberg, P. K., and M. M. Knetter. 1997. "Goods Prices and Exchange Rates: What Have We Learned?" *Journal of Economic Literature* 35 (3): 1243 72. A useful and much cited literature review on the exchange rate pass through with particular reference to the United States.

Knetter, M. M. 1993. "International Comparisons of Pri cing to Market Behavior." *American Economic Review* 83 (3): 473 86. One of the earliest papers introducing the concept of pricing to market.

Taylor, J. B. 2000. "Low Inflation, Pass Through, and the Pricing Power of Firms." *European Economic Review* 44 (7): 1389 1408. The pioneering study linking stable monetary policy with the low and declining exchange rate pass through in industrial countries.

AMIT GHOSH AND RAMKISHEN S. RAJAN

■ exchange rate regimes

The exchange rate regime is the characterization of the way a government manages its national currency in the foreign exchange market. The different forms of exchange rate regimes relate to how actively the government manages the foreign exchange rate, and how its actions are supported by institutional arrangements. At one end of the spectrum is an independently floating currency; at the other, a fixed exchange rate system, sometimes referred to as a "hard peg." Many varieties of exchange rate regimes exist between these two poles.

Floating Rate Regimes With an independent floating exchange rate regime the exchange rate is market determined and there is no explicit exchange rate target level or institutional commitment to influence the path of exchange rates. The purest form of independent floating is the "free float," in which the government neither intervenes in the foreign exchange market (buying or selling foreign currency from international reserves) nor directs monetary, fiscal, or regulatory policies to an explicit exchange rate objective. In free float regimes, the government allows the market-determined exchange rate to prevail and does not attempt to influence its value. Instead, the government directs macroeconomic policy entirely toward domestic stabilization objectives such as output and inflation levels.

Few countries follow purely free floats; even those that are independently floating occasionally intervene in foreign exchange markets and take into account the value of the exchange rate in making macroeconomic policy decisions. The countries with regimes most closely approaching pure free floats are the United States and the euro area (European countries that share the euro as a common currency). The United States has intervened in the foreign exchange market on only a few occasions since the mid-1990s with the objective of countering "disorderly" market conditions. The Federal Reserve System also puts little weight on exchange rates in its policy deliberations. Similarly, the European Central Bank has intervened in the foreign exchange market only a few times since the introduction of the euro in 1999 (Fatum and Hutchison 2002).

Other countries with independently but not freely floating regimes include Japan, Canada, New Zealand, Sweden, and many more. Countries that follow independent floats intervene in foreign exchange markets occasionally and at times quite heavily, but have no explicit exchange rate target level. Japan, for example, intervened to an unprecedented degree in 2003 and 2004, buying more than $300 billion in the foreign exchange market in an attempt to support the value of the U.S. dollar

exchange rate against the yen (Fatum and Hutchison 2006).

Closely related to independently floating systems are "managed floating" exchange rate regimes. In these cases, the exchange rate is again largely market determined, with official foreign exchange market intervention aimed at slowing exchange rate movements or reducing volatility but without having a specific exchange rate path or target. These countries tend to undertake more active interventions in foreign exchange markets than independent floaters, with indicators for managing the rate being broadly judgmental (e.g., balance of payments position, international reserves, parallel market developments), and adjustments may not be automatic. In managed floating cases, however, the exchange rate is typically a secondary policy target with little constraint placed on the conduct of domestic monetary or fiscal policy. The International Monetary Fund (IMF) characterized 53 exchange rate regimes at the end of 2005 as "managed floating, with no predetermined path for the exchange rate." Most developing and emerging market economies fall within the managed floating group. Examples include the Czech Republic, Paraguay, and Indonesia.

National Legal Tender, Monetary Unions, and Currency Boards A "fixed" exchange rate regime is at the opposite end of the spectrum from an independent floating regime. In a fixed exchange rate system, the government is committed to maintaining a particular exchange rate value and pegs this rate by a combination of official foreign exchange market intervention, macroeconomic policies, and institutional arrangements such as international capital controls. The "hardest," or most durable and credible, exchange rate peg is supported by strong institutional arrangements, either by stringent controls on the flows of financial capital into and out of the country, when the country enters into a currency board arrangement, or when the national currency is replaced with a foreign currency ("dollarization").

Dollarization refers to a country's choice to circulate the currency of another country as its sole legal tender, without a separate national currency. The term *dollarization* does not necessarily refer to the

U.S. dollar; it is a generic term for any national currency that is the legal tender in a foreign country. Ecuador "dollarized" after currency turmoil, and the U.S. dollar has been in circulation in Panama for many years. Many small island nations are dollarized because it would be impractical for them to have independent national currencies. The IMF classified 29 countries (excepting members of the European Monetary Union) in the group "exchange rate arrangements with no separate legal tender" at the end of 2005. Dollarized countries have no control over the exchange rate, and typically little or no control over domestic monetary policy, since interest rates are largely determined abroad.

Countries that belong to monetary or currency unions have a legal tender that is shared by members of the union. As of 2006, 12 European countries had given up their national currencies since 1999, adopting the euro as a common currency. Previously separate (except in the case of Belgium and Luxembourg) national legal tenders have been replaced by one currency and one exchange rate, which is independently floating. Monetary unions have a common monetary and exchange rate policy. In the case of the euro, monetary policy is run by the European Central Bank (ECB) in Frankfurt. Unlike dollarized countries, however, countries in modern monetary unions participate in the exchange rate and monetary policy deliberations for their shared currency. The ECB has a decision-making structure representing all of the European Monetary Union (EMU) countries.

The "hardest" and most credible of the pegged rate systems (considering arrangements with no separate legal tender as a special case) is the currency board regime. Under this arrangement, all of the national currency in circulation is backed by the government's holding of foreign currency reserves, and the government stands ready to buy or sell as much foreign currency as necessary to maintain the exchange rate value. Currency boards are based on an explicit legislative commitment to exchange domestic currency for foreign currency at a fixed exchange rate. To enable the central bank to make this commitment, the board imposes restrictions on the is-

suance of domestic currency: new issuance of national legal tender is contingent on an inflow of foreign exchange reserves. The government has little or no control over monetary policy or short-term interest rates in a currency board institutional setting. The Hong Kong Special Administrative Region, China, has a long-standing currency board arrangement. Estonia, Lithuania, and several other countries also have currency board arrangements.

Pegged Exchange Rate Arrangements Pegged exchange regimes may take a variety of forms, including fixed peg arrangements, pegged rates in horizontal bands (e.g., the Danish krone vis-à-vis the euro), crawling pegs (e.g., Turkey), rates within crawling bands (e.g., Hungary, Israel, and Venezuela), and other fixed pegs (e.g., China, Egypt, and Malaysia). In a conventional fixed peg regime, the authorities attempt to maintain the value of the exchange rate within a narrow band (usually plus or minus 1 percent) around a central parity (or reference value) exchange rate. The authorities commit to intervene by buying and selling foreign exchange reserves, sometimes in cooperative arrangements with other central banks, or to adjust interest rates and other policies to maintain the exchange rate within the band. There is no legal or institutional commitment to maintain the central parity exchange rate irrevocably, and occasional adjustments to the central parity may occur. The central parity referenced exchange rate may be a single currency, a cooperative arrangement, or a currency composite. In the absence of capital controls, there is very limited monetary autonomy with a fixed rate peg, as foreign exchange interventions, and often interest rate changes, are necessary to maintain the value of the currency within a narrow range. Forty-five countries followed a fixed peg arrangement at the end of 2005, reflecting a wide range in level of development and geographic distribution. Examples include China, Malaysia, Nepal, and Saudi Arabia.

A pegged rate within horizontal bands is similar to a fixed peg but with more flexibility of the exchange rate around the central parity, typically plus or minus 2 percent. The wider band, compared to the fixed peg regime, allows somewhat greater monetary policy discretion. In general, the broader the band width, the greater the degree of monetary policy independence in principle, monetary policy discretion is possible until the limits of the band are reached. As in the case of conventional fixed pegs, reference may be made to a single currency, a cooperative arrangement, or a currency composite. Target zones, such as the former exchange rate mechanism (ERM) of the European Monetary System, are a form of pegged rates within a horizontal band. Only six countries followed a pegged rate within horizontal band arrangement at the end of 2005, four of which were in the context of the ERM II arrangement with the EMU (Cyprus, Denmark, the Slovak Republic, and Slovenia).

Two other common forms of pegged exchange rate regime are the crawling peg and crawling band systems. The crawling peg refers to a case in which the currency is adjusted periodically in small amounts, either at a fixed increment or in response to indicators such as domestic inflation relative to trading partners or perhaps to expected future inflation differentials. Maintaining a crawling peg imposes constraints on monetary policy in a manner similar to a fixed peg system. Similarly, the crawling band regime maintains the currency within a band but adjusts the central reference rate or margins periodically in response to quantitative indicators.

Exchange Rate Regimes, Monetary Policy, and International Capital Controls One key feature distinguishing exchange rate regimes is the extent to which they constrain the conduct of monetary policy. Generally speaking, the greater the fixity of the exchange rate, the less room there is for monetary autonomy. On one side of the spectrum is an independently floating exchange rate, in which conduct of domestic monetary policy is under little constraint. On the other side of the spectrum is a currency board arrangement or a country that has no separate national legal tender. In these latter two cases, there is little or no room for independent monetary policy. If there is no separate legal tender, there is no national money and therefore no national monetary policy. In the currency board arrangement, changes in monetary policy follow the balance of

payments and the rise and fall of official international reserves. Capital outflows that result in a decline of international reserves automatically trigger a contraction of the money supply, and the opposite is the case for capital inflows.

As the exchange rate regime moves along the spectrum from hard pegs toward independently floating, monetary authorities in principle have greater control over setting domestic interest rates and the path of monetary policy. The degree of international capital mobility, that is, the legal, institutional, or market-related constraints on the cross-border flow of financial resources, also affects a country's monetary policy. Monetary authorities frequently impose restrictions on capital flows in an attempt to gain more monetary autonomy in the face of constraints imposed, say, by fixed pegged exchange rates. The Bretton Woods system of fixed peg exchange rates (1946 71) lasted several decades because of pervasive capital controls, policies that led to low inflation in the United States, and the willingness of countries pegged to the U.S. dollar to accept limited monetary autonomy. Countries are more willing to give up monetary independence and link with the U.S. dollar if the monetary leadership of the reserve-currency country provides a stable macroeconomic environment.

The economics term *impossible trinity* means that countries cannot simultaneously follow an independent monetary policy, maintain a fixed peg exchange rate regime, and allow unrestricted international capital flows. Countries frequently face a dilemma if their domestic stabilization policy objectives and desired path of monetary policy (level of interest rates, for example) conflict with the monetary policy needed to maintain a pegged exchange rate. In these circumstances the country must abandon its desire to pursue its stabilization objectives, give up the exchange rate peg, or impose capital controls in the hope of gaining some flexibility in reconciling the two other objectives. Exchange rate regime changes frequently occur because of this dilemma, usually because countries cannot constrain monetary and fiscal policy effectively to curtail inflation and balance of payments imbalances. In a

system with freely mobile international capital, this policy dilemma is often exacerbated by a sharp capital outflow as domestic asset holders, both foreign and national residents, anticipate devaluation (and associated capital loss on domestic assets) and move to sell the domestic currency before it occurs. In this circumstance, devaluation of the exchange rate is commonplace.

Optimal Exchange Rate Regimes and Regime Collapses A basic question in discussing exchange rate regimes concerns which regime is best suited to the particular circumstances of a given country. This topic is often referred to as "optimal" exchange rate regime choice. Much of the work in this area focuses on the nature of disturbances facing the country, whether they be foreign or domestic in origin, real or monetary, permanent or transitory, combined with the choice made by authorities of a particular systematic approach to intervention in the foreign exchange market or monetary policy rule to influence exchange rate developments (Marston 1985). The exchange rate arrangement must be consistent with the degree of monetary independence and international capital mobility to be sustainable. Traditional theory suggests that an intermediate regime usually is preferable to either a hard peg, such as a currency union or currency board arrangement, or a completely free floating exchange rate. The structure of the economy, the nature of economic disturbances, and the degree of capital mobility are some of the standard criteria in the choice of an exchange rate regime.

The choice of exchange rate regime, however, is more constrained than the traditional view would suggest. Speculative international capital flows can cause regime collapses very quickly, even when current fundamentals appear strong. Concerns about future policy choices, vulnerabilities in the banking sector, and similar factors may start a run on the currency. Self-fulfilling crises are also possible, with a seemingly random speculative attack causing so much economic disruption that the policy authorities may change course entirely, devalue the currency, and end up with higher inflation than otherwise would be the case. Clearly, exchange rate regime

sustainability and choice are greatly complicated in an era of integrated financial markets and deregulated capital flows, and in response there has been a tendency for countries to move to one polar case or the other either hard pegs or independently floating exchange rates.

See also band, basket, and crawl (BBC); Bretton Woods system; capital controls; capital mobility; common currency; currency board arrangement (CBA); currency crisis; currency substitution and dollarization; discipline; equilibrium exchange rate; European Central Bank; European Monetary Union; exchange rate forecasting; exchange rate volatility; Federal Reserve Board; financial crisis; foreign exchange intervention; hot money and sudden stops; impossible trinity; inflation targeting; International Monetary Fund (IMF); international reserves; monetary conditions index; purchasing power parity; real exchange rate; sterilization

FURTHER READING

Eichengreen, Barry, and Charles Wyplosz. 1993. "The Unstable EMS." *Brookings Papers on Economic Activity* 1. Washington, DC: Brookings Institution. Investigates the 1992 EMS crisis, offering four explanations and emphasizing the constraints of the "impossible trinity" on capital flows, fixed exchange rates, and monetary policy.

Fatum, Rasmus, and Michael Hutchison. 2002. "ECB Foreign Exchange Intervention and the Euro: Institutional Framework, News, and Intervention." *Open Economies Review* 13 (4) (October): 412 25. Explains the foreign exchange market policy of the newly established European Central Bank and the effect of reported intervention on the euro exchange rate.

———. 2006. "Effectiveness of Official Daily Foreign Exchange Market Intervention Operations in Japan." *Journal of International Money and Finance* 25: 199 219. Measures the effectiveness of large scale intervention operations by the Bank of Japan using an event study approach with daily data.

Fischer, Stanley. 2001. "Exchange Rate Regimes: Is the Bipolar View Correct?" *Journal of Economic Perspectives* 15 (2) (spring): 3 24. Presents a broad survey of the history and theory of exchange rate regime failures and explores whether hard currency pegs or flexible rates are

the only practical alternatives for exchange rate regime choice.

International Monetary Fund. 2006. "De Facto Classification of Exchange Rate Regimes and Monetary Policy Framework, as of December 31, 2005." Downloadable from http://www.imf.org/external/np/mfd/er/index.asp. Presents data on the classification of exchange rate regimes and monetary policy frameworks for IMF member countries.

Krugman, Paul. 1979. "A Model of Balance of Payments Crises." *Journal of Money, Credit, and Banking* (August): 311 25. A classic, which shows how fundamental macroeconomic imbalances lead to predicable runs on central bank reserves.

Marston, Richard C. 1985. "Stabilization Policies." In *Handbook of International Economics*, vol. 2, edited by R. W. Jones, P. B. Kenen, G. M. Grossman, and K. Rogoff. Amsterdam: Elsevier, 859 916. A survey of macroeconomic policies in open economies with a section on the optimal exchange rate regime choice as a function of underlying disturbances, structure of the economy, and policy objective function.

Obstfeld, Maurice. 1996. "Models of Currency Crisis with Self Fulfilling Features." *European Economic Review* 40 (April): 1037 47. A classic contribution demonstrating that currency crises may be associated with self fulfilling speculative attacks and showing the existence of multiple equilibrium exchange rates when the policy authorities and private sector behave strategically.

MICHAEL M. HUTCHISON

■ exchange rate volatility

One of the key facts in international finance is that when exchange rates are flexible they are also highly volatile. Such volatility manifests itself in two key ways: inter- and intraregime volatility. *Intraregime volatility* refers to the volatility of exchange rates in floating exchange rate regimes relative to the volatility of macroeconomic fundamentals such as money supplies, interest rates, and output levels. A common, although by no means universal, belief is that exchange rates are excessively volatile compared to macroeconomic fundamentals. *Interregime vola-*

tility refers to the behavior of real and nominal exchange rates in the move from fixed to floating exchange rates: as a country moves from fixed to floating exchange rates the volatility of standard macroeconomic fundamentals remains unchanged, but the volatility of both real and nominal exchange rates increases dramatically. The key question that arises from this kind of regime-dependent behavior is: Can such exchange rate behavior be explained in a manner consistent with a macroeconomic framework (that is, using a model that relies on macroeconomic fundamentals such as money supplies), or does such a framework have to be abandoned and an alternative framework, such as a market microstructure approach, be adopted?

The first part of this entry analyzes the issue of intraregime volatility and, specifically, different explanations for the volatility result. All the explanations rely on some variant of the monetary model of the exchange rate. The second part of the entry focuses on interregime volatility, starting with empirical evidence and then considering some theoretical explanations for this phenomenon.

Intraregime Volatility: Monetary Model In trying to explain intraregime volatility in terms of macroeconomic fundamentals, the monetary model of the exchange rate has become something of a workhorse. In essence this model posits that exchange rates are determined by the interaction of money supply (usually assumed to be exogenous, i.e., predetermined by the central bank) and money demand (which is a function of interest rates and income levels) terms between the home country and its trading partners. Other things being equal, an increase in the home money supply leads to a proportionate depreciation in the exchange rate.

In the flexible price version of the monetary model, exchange rate volatility is explained in terms of a "magnification effect" that arises because of the link in the model between the current (spot) exchange rate and the future expected exchange rate (Bilson 1978). The model may be written simply as:

$$s_t = f_t + E_t s_{t+1}, \qquad (1)$$

where s_t denotes the nominal exchange rate, f_t is a composite fundamental comprising home and for-

eign money supplies and income terms (suitably weighted with income elasticities), and $E_t s_{t+1}$ is the expected exchange rate in period t+1. The expected exchange rate in each future period of the life of the underlying asset (money) is determined by expected money supplies and expected income levels and so the current spot exchange rate in this model becomes a function of current fundamentals (money and income) in other words, period-t variables and the expected fundamentals in all future periods. More formally, the current spot price is the present discounted value of all expected fundamentals. To the extent that a current change in the money supply signals to agents an expected increase in future period fundamentals, this can cause the exchange rate to move by more than the current change in the period-t fundamentals, f_t. By simply observing the current exchange rate and the current change in fundamentals, a magnified response of the former with respect to the latter will be seen. This is the so-called magnification effect.

The monetary model has also been used to demonstrate the implications that speculative bubbles or, more generally, any nonfundamental factor can have for exchange rate volatility. A speculative bubble or nonfundamental factor can be added to equation (1) as:

$$s_t = f_t + E_t s_{t+1} + b_t, \qquad (2)$$

where b_t represents a speculative bubble term and can impart exchange rate volatility over and above that generated by the macroeconomic fundamentals. The speculative bubble can be a rational bubble, which means it is consistent with the underlying model (i.e., equation 1), or nonrational and therefore not necessarily consistent with any model. The introduction of a speculative term into the monetary model can generate excessive volatility of the exchange rate with respect to the fundamentals (see MacDonald 2007).

Intraregime Volatility: Dornbusch Overshooting Model Perhaps one of the best-known explanations for intraregime volatility is the seminal overshooting model of the economist Dornbusch (1976). This model is also in the monetary class, but assumes that consumer prices are sticky in the short run (although flexible in the long run) while asset

prices and yields (the exchange rate and interest rates) are continuously flexible. In the short run, income is assumed to be fixed. In this context, an increase in the domestic money supply upsets money market equilibrium and leads to a proportional depreciation of the long-run, or equilibrium, price level and exchange rate (which are flexible). In the short run, however, or in the immediate aftermath of the money supply increase, equilibrium can be restored only by a change in the interest rate (since the price level and income are held constant by assumption). But in a world of high, or complete, capital mobility the domestic interest rate is tied into the foreign rate through the uncovered interest rate parity condition and can change only if the expected change in the exchange rate is assumed to be nonzero.

The expected change in the exchange rate is, however, nonzero in the overshooting model because it is governed by a regressive expectations mechanism: a current depreciation of the exchange rate relative to its equilibrium, or long-run, value is expected to be reversed in the future. In other words, a current depreciation of the exchange rate produces the expectation of a future appreciation. So, in response to the increase in the money supply, what is required is the current exchange rate to move more than proportionally to the long-run, or equilibrium, exchange rate value (which moves in proportion to the increase in the money supply) to allow the domestic interest rate to fall below the world level: the exchange rate overshoots its long-run value and, by implication, the current change in the money supply. As in the magnification story, the current exchange rate is more volatile than current fundamentals. The extent of the overshooting is governed by how sensitive interest rates are to money supply changes (the less sensitive, the more the exchange rate will move) and the sensitivity of the expected change in the exchange rate with respect to the gap between the current and equilibrium exchange rates.

Intraregime Volatility: NOEM The variant of the monetary model used to derive both the magnification and overshooting results is regarded as ad hoc, in the sense that it is not based on an explicit microfoundations approach   that is, the starting point is

not an explicit utility function, maximized subject to a budget constraint. Both Stockman (1980) and Lucas (1982), however, present variants of the monetary model founded on an explicitly optimizing framework in which agents maximize utility subject to a budget constraint and a cash-in-advance constraint. In this model the current exchange rate is, as in the ad hoc approach, a function of relative money supplies and output levels and also the ratio of home to foreign marginal utility. In this context exchange rate volatility with respect to current information (in this case, period-t relative prices) arises because both preference and technology shocks   real shocks   are highly volatile. More recent explanations for exchange rate volatility using an optimizing approach have exploited the New Open Economy Macroeconomics (NOEM) model of Obstfeld and Rogoff (1995) and we now provide a brief summary of these explanations.

By introducing uncertainty into the basic NOEM, Obstfeld and Rogoff (2000) generate a variant of the forward-looking monetary model in which a "level" risk premium term features. There are two key insights here. First, the risk premium can affect the level of the exchange rate, and not just the predictable excess return, which has been studied extensively in the literature (see, e.g., MacDonald 2007). This is important because it means that higher moments of economic variables can affect the volatility of the exchange rate and not just the first moment   if the forward risk premium is quite volatile, this could have important implications for exchange rate volatility. Second, the effect of the risk premium on the exchange rate may potentially be very large in this model because it enters the forward-looking model with a scaling factor that puts a large weight on the premium.

In the Obstfeld and Rogoff (2000) model a rise in the level of home monetary variability leads to a fall in both the level exchange risk premium and the forward exchange rate risk premium, and, given plausible values of the scaling factor, imparts an exchange rate appreciation and considerable exchange rate volatility. The appreciation of the exchange rate and fall in the risk premium are contrary to the standard

intuition for the effect of increased monetary variability in which the common casual presumption is that financial markets attach a positive risk premium to the currency of a country with high monetary volatility. The effect is different in this model because in the sticky price variant, positive monetary shocks lead to increases in global consumption, which means that domestic money can be a hedge, in real terms, against shocks to consumption. Furthermore, higher monetary variability raises the expectation of the future real value of money (other things being equal) which is the convexity term (this effect also works in a flexible price model).

A further attempt to explain the excess volatility of exchange rates using a variant of the NOEM is made by Devereux and Engel (2001). They attempt to shed light on a conjecture of Krugman (1989) that exchange rate volatility is so great because fluctuations in the exchange rate matter so little for the economy. They use a variant of the NOEM in which there is a combination of local-currency pricing, heterogeneity in international price setting and in the distribution of goods (for example, some firms market their products directly in the foreign market and charge a foreign price while some exporters use foreign distributors, charging a price set in the exporter's currency), and crucially, the existence of noise traders (that is, traders who do not base their trades on fundamentals) who impart expectational biases into international financial markets.

Devereux and Engel (2001) demonstrate that in this setup the conditional volatility of the exchange rate depends on the volatility of the fundamentals (which in their case is the volatility in relative money supply terms) and the extent of local-currency pricing. With complete local-currency pricing, the conditional volatility of the exchange rate effectively rises without bound. This is because the combination of local-currency pricing, along with asymmetric distribution of goods and noise trading, implies a degree of exchange rate volatility that is far in excess of the underlying fundamental shocks.

Intraregime Volatility: Foreign Exchange Microstructure All of the foregoing research assumes that expectations are homogeneous. Bacchetta and

Wincoop (2003) build on the idea that the heterogeneity of investors may also be important for an understanding of exchange rate dynamics, however. In particular, they introduce two types of investor heterogeneity that have been associated with order flow into a standard variant of the monetary model. The first type is the heterogeneous information of market participants about future macro fundamentals a dispersion effect and the second is heterogeneity due to nonfundamentals. The latter includes noise traders and rational traders who trade for nonspeculative reasons, such as liquidity trades, or trades associated with differential access to private investment opportunities. Bacchetta and Wincoop demonstrate how information heterogeneity produces both a magnification effect on the exchange rate and endogenous persistence of the impact of nonfundamentals on the exchange rate. Using a simulation exercise and plausible parameter values, the authors demonstrate that there is a substantial magnification effect as a result of information dispersion, and a substantial part of this seems to be attributable to the role of higher-order expectations due to the infinite regress.

A number of economists have attempted to gauge the forward-looking monetary model as written in (1) and its speculative bubble variant given in (2) (e.g., see MacDonald and Taylor 1993). The evidence, in general, does not support the magnification story, although this could simply reflect the model specifications used. The research gives some support to the speculative bubbles hypothesis, although such tests could simply be picking up some other extraneous influence on the exchange rate (see, e.g., Engel and West 2004). More favorable empirical evidence has been reported for the overshooting model, while tests of the more recent NOEM variants of the monetary model have not yet begun. So, at best, we can conclude from this empirical evidence that the jury is still out on the validity of the various theoretical explanations for intraregime volatility (see also Arnold, MacDonald, and de Vries 2005).

Interregime Volatility Baxter and Stockman (1989) were the first to examine the variability of output, trade variables, and private and government

consumption across the Bretton Woods and post Bretton Woods experience, and they demonstrated that the behavior of macroeconomic fundamentals does not change in the move from fixed to floating exchange rates but that the behavior of the exchange rate (both real and nominal) does.

Flood and Rose (1995) reexamined the issue of interregime volatility using the monetary model. Specifically, they construct composite f terms using standard monetary and income variables and compare the conditional volatility of this term to the conditional volatility of the sum of the exchange rate and the interest differential for the Bretton Woods and post Bretton Woods periods. The countries studied are the United Kingdom, Canada, France, Germany, Holland, Italy, Japan, and Sweden. The results of Flood and Rose show that the volatility of the exchange rate term increases dramatically as countries move from fixed to floating, but the volatility of the composite fundamental term does not. In their view, the standard monetary fundamentals cannot explain the volatility of exchange rates in the move from fixed to floating exchange rates and attention has to be focused on what does change in the fixed to floating move, namely the market microstructure of the foreign exchange market.

A number of papers have attempted to understand the interregime puzzle without abandoning a macroeconomics-based approach. For example, Duarte (2003) uses a variant of the two-country NOEM model in which asset markets are incomplete and prices are set one period in advance in the buyer's currency (i.e., local-currency pricing) to address the interregime volatility issue. Duarte studies the properties of this model in the context of a simulation exercise in which the utility function is fully specified, along with technology and monetary shocks. This exercise clearly generates a sharp increase in the volatility of the real exchange rate following a switch from fixed to flexible rates, with no similar change in the volatilities of output, consumption, or trade flows. The intuition for this result is quite simple: because prices are set one period ahead in the buyer's currency, allocation decisions are disconnected at the time of impact from unexpected changes in the nominal exchange rate and so the volatilities of output, consumption, and trade flows are unaffected.

Arnold, MacDonald, and de Vries (2005) argue that to understand interregime volatility issues it is important to introduce distortions (such as trade and capital account restrictions and IMF support), which are likely to be prevalent in fixed rate regimes such as Bretton Woods, into the standard monetary exchange rate equations. Using a variety of measures for the distortion terms, the authors demonstrate that such terms can explain the interregime puzzle.

Reinhart and Rogoff (2002) argue that the interregime puzzle may not be such a puzzle after all. Specifically, they argue that empirical studies such as those of Flood and Rose (1995) and Baxter and Stockman (1989) use the IMF method of classifying an exchange rate regime as fixed or flexible. The de facto situation in the Bretton Woods and post Bretton Woods eras, however, was in fact very different, with the majority of countries in Bretton Woods having much more flexibility in their exchange rate behavior than that portrayed in the official statistics, and in the floating rate period there is in practice much more evidence of exchange rate fixity than that indicated in the standard classification. Perhaps the neatest explanation for interregime volatility is that when exchange rates are de jure fixed they are often de facto flexible, and when they are flexible they are in practice fixed or quasi-fixed.

See also balance of payments; Bretton Woods system; exchange rate forecasting; exchange rate pass-through; exchange rate regimes; foreign exchange intervention; interest parity conditions; money supply; New Open Economy Macroeconomics; peso problem; quantity theory of money; real exchange rate; speculation; sterilization

FURTHER READING

Arnold, I., R. MacDonald, and C. de Vries. 2005. "Fundamental Volatility Is Regime Specific." Mimeo. University of Glasgow, Scotland. Revisits the evidence on intraregime exchange rate volatility.

Bacchetta, P., and E. van Wincoop. 2003. "Can Information Heterogeneity Explain the Exchange Rate Determination Puzzle?" NBER Working Paper No. 9498. Cambridge, MA: National Bureau of Economic Re

search. Emphasizes the importance of investor hetero geneity in understanding exchange rate dynamics.

Baxter, M., and A. Stockman. 1989. "Business Cycles and the Exchange Rate System." *Journal of Monetary Economics* 23: 377 400. Empirically investigates the differences in time series behavior of output, consump tion, trade flows, government consumption spending, and real exchange rates under alternative exchange rate systems.

Bilson, John F. O. 1978. "Rational Expectations and the Exchange Rate." In *The Economics of Exchange Rates: Selected Studies*, edited by Jacob A. Frenkel and Harry G. Johnson. Reading, MA: Addison Wesley, 75 96. Clas sic paper on the flexible price version of the monetary model of exchange rate determination.

Devereux, M., and C. Engel. 2001. "Endogenous Currency Price Setting in a Dynamic Open Economy Model." Mimeo. University of Wisconson Madison. Well cited paper on the NOEM's rationalization of excess volatility of exchange rates.

Dornbusch, R. 1976. "Expectations and Exchange Rate Dynamics." *Journal of Political Economy* 84: 1161 76. Classic paper on the role of expectations and exchange rate overshooting.

Duarte, M. 2003. "Why Don't Macroeconomic Quantities Respond to Exchange Rate Variability?" *Journal of Monetary Economics* 50: 889 913. Examines why ex change rate movements do not respond to macroeco nomic variables.

Engel, C., and K. D. West. 2004. "Accounting for Exchange Rate Variability in Present Value Models When the Discount Factor Is Near One." NBER Working Paper No. 10267. Cambridge, MA: National Bureau of Eco nomic Research. Investigates why nominal exchange rates in low inflation advanced countries are almost random walks.

Flood, R. P., and A. Rose. 1995. "Fixing Exchange Rates: A Virtual Quest for Fundamentals." *Journal of Monetary Economics* 36: 3 37. Revisits interregime exchange rate volatility.

Krugman, P. 1989. *Exchange Rate Instability*. Cambridge, MA: MIT Press. Highly readable monograph on the extreme volatility of exchange rates in the 1980s.

Lucas, R. E. 1982. "Interest Rates and Currency Prices in a Two Country World." *Journal of Monetary Economics*

10: 335 60. Presents variants of the monetary model founded on an explicitly optimizing framework.

MacDonald, R. 2007. *Exchange Rate Economics: Theories and Evidence*. London: Taylor Francis. Offers an up to date and comprehensive overview of exchange rate measurement and determination.

MacDonald, R., and M. P. Taylor. 1993. "The Monetary Approach." *IMF Staff Papers* 40: 89 107. Reviews the monetary approach to exchange rate determination.

Monacelli, T. 2004. "Into the Mussa Puzzle: Monetary Policy Regimes and the Real Exchange Rate in a Small Open Economy." *Journal of International Economics* 62: 191 217. Revisits the issue of interregime exchange rate volatility.

Obstfeld, M., and K. Rogoff. 1995. "Exchange Rate Dyna mics Redux." *Journal of Political Economy* 103: 624 60. ———. 2000. "Risk and Exchange Rates." NBER Work ing Paper No. 6694. Cambridge, MA: National Bureau of Economic Research. Two well cited papers on the NOEM's rationalization of excess volatility of exchange rates.

Reinhart, C. M., and K. S. Rogoff. 2002. "The Modern History of Exchange Rate Arrangements: A Rein terpretation." NBER Working Paper No. 8963. Cam bridge, MA: National Bureau of Economic Research. Well cited paper that develops a new system of classi fying historical exchange rate regimes.

Stockman, A. 1980. "A Theory of Exchange Rate De termination." *Journal of Political Economy* 103: 673 98. Present variants of the monetary model founded on an explicitly optimizing framework.

RONALD MACDONALD

■ exchange rate weapon

The concept of the *exchange rate weapon* refers to the use of the exchange rate by one state to secure policy change on the part of another. The weapon helps to explain important aspects of the international polit ical economy of monetary relations, such as the be havior of governments and central banks during adjustment conflicts and the distribution of the costs of adjustment among countries. When collective management of the adjustment process has been re-

quired, the use or threat of use of the exchange rate weapon often underpinned international agreements.

The exchange rate weapon can take two forms, one *passive* and the other *active*. In the passive form, officials of a leading state within the system can allow the exchange rate of its currency to shift, perhaps even overshoot, in the knowledge that a more vulnerable partner country is thus subject to incentives to adjust its fiscal and/or monetary policy. Authorities can also actively promote a shift in the rate to induce the partner to shift macroeconomic policy. The two forms are deliberate, often coincide empirically, and have similar coercive effects.

Adjustment Conflict and Causal Mechanism
The exchange rate weapon can come into play particularly when current account imbalances become unsustainable and key states enter into political conflict over measures to reduce them. A large, closed economy is less sensitive to changes in the exchange rate compared to its smaller, more open partner(s), other things being equal. Under these circumstances, the smaller state will be subject to greater pressures to change macroeconomic policies in the face of a large shift in the exchange rate. The presence of a substantial asymmetry enables the large country to alter the incentives that confront the small country when setting macroeconomic policy.

Consider the example of a large country with a current account deficit locked in an adjustment standoff with a small partner, which runs corresponding surpluses. Domestic politics in both countries have dictated the macroeconomic policy preferences (e.g., for employment, growth, and inflation) that have given rise to the imbalances. The large state could eliminate the imbalance by shrinking its government budget deficit, for example, and the smaller state could accomplish the same through a fiscal expansion. For domestic political reasons, however, each government prefers that the other undertake the policy change.

When the imbalance becomes unsustainable that is, private markets have become unwilling to finance it the governments are confronted with politically difficult choices. The large, less vulnerable state could allow or encourage its currency to depreciate with relative equanimity. The corresponding currency appreciation of the small, more vulnerable state, however, sets in motion a chain of events that alters that government's incentives with respect to its macroeconomic policy. Specifically, the appreciation of the small country's currency will reduce the domestic price of imports, reducing inflation, and will shift the current account balance toward deficit, reducing growth and employment prospects. The reduction in inflation eases a constraint on stimulating domestic demand and the reduction in growth provides an incentive to do so. A change in macroeconomic conditions within the small country thus shifts the domestic political equilibrium toward providing a stimulus.

An asymmetry in exchange rate vulnerability enables the more insulated states to exploit cross-border economic effects in applying political pressure on their partners for policy adjustment. This generalization holds across a broad range of open economy macroeconomic assumptions, such as the degree of capital mobility. Three additional factors can reinforce the strength of the exchange rate weapon. First, a large, less vulnerable country might press the small country for policy change through diplomatic moral suasion, offering its support for currency stabilization in return. Second, the large country could deliberately act to shift the exchange rate in order to secure policy change on the part of the small country. Third, the large country could be reinforced in this effort by the international role of its currency, as in the case of the United States.

The active use of the exchange rate weapon hinges on the "exogeneity" of the exchange rate that is, the ability of governments to influence currency values without altering underlying macroeconomic policies. Governments' scope for doing so is widely debated among international monetary economists. Professional consensus on the question has evolved over the decades. Recent studies generally find intervention to be more effective than did studies conducted during the 1980s. Suffice it to say that government action can be successful under a variety of circumstances, such as when it is publicly an-

nounced, conducted jointly by two or more central banks, and consistent with the underlying fundamentals, and when the exchange rate is far from equilibrium.

Historical Episodes The exchange rate weapon became a threat or a tactic during each of four episodes of conflict over international adjustment since the Bretton Woods regime: (1) the breakdown of that regime in the early 1970s, (2) the Bonn Summit of 1978, (3) the Plaza and Louvre accords of the mid-1980s, and (4) the early and mid-1990s. During each of these episodes, U.S. administrations pressed European and Japanese governments and/or central banks for expansionary measures and in some cases actively encouraged a depreciation of the dollar. U.S. authorities were more successful in some episodes, such as 1971–73, 1977–78, and 1985–87, than in others, such as during the 1990s.

Consider the Plaza-Louvre period as an example. During the first half of the 1980s, the U.S. current account deficit increased substantially under the macroeconomic and exchange rate policies of the first Reagan administration. In the summer of 1985, political pressure for trade protection in the U.S. Congress boiled over and Treasury Secretary James A. Baker III sought to defuse it with international policy coordination, among other measures. He and his counterparts among the Group of Five finance ministers signaled their support for dollar depreciation in the Plaza Accord of September of that year. A few months later, but prior to administering substantial stimuli to their economies, Japanese and West German officials indicated that the dollar had depreciated enough. By contrast, U.S. Treasury officials continued to "talk down" the dollar further and very large appreciations of the yen and mark occurred, prompting lobbying within Japan and West Germany for expansionary measures to offset the contractionary effects. At the meeting at the Louvre in 1987, U.S. officials agreed to stabilize currencies in exchange for European and Japanese policy stimuli. Changes in exchange rates and macroeconomic policy helped to nearly eliminate the U.S. current account deficit in subsequent years.

Structural Shift Because the effective use of this weapon hinges on the exploitation of an asymmetry by a large state, changes in the relative size and openness of the major countries could well affect the relevance of the weapon during adjustment conflicts in the future. Changes in the broad structure of the global economy structural shift derive from three main developments. First, the formation of the euro area in January 1999 created, in one fell swoop, a new monetary region of roughly equivalent weight to the United States the most important change in the structure of the international monetary system since at least the Bretton Woods regime. Second, in East Asia the adjustment problem has been regionalized by foreign direct investment and cross-border production networks, as well as the practice by governments within the region of shadowing Chinese exchange rate policy vis-à-vis the dollar in order to preserve trade competitiveness. East Asian governments are advancing potentially important projects of regional financial cooperation. Finally, the U.S. economy is increasingly internationalized as measured by trade relative to gross domestic product and foreign investment relative to total U.S. financial assets.

The combined effect of these developments is to make Europe potentially less vulnerable to dollar depreciation, give pause to American officials when contemplating use of the exchange rate weapon in East Asia, and reduce the asymmetry in the effects of exchange rate changes between the United States and its partners. The exchange rate weapon could therefore be less useful to U.S. policymakers in the future than it has been in the past. Nonetheless, the United States will probably retain residual power while Europe's monetary union, Japan, and China could conceivably begin to exercise a similar form of influence through their currencies in geographically contiguous areas. The exchange rate weapon is thus likely to have continuing relevance as an instrument of economic conflict.

Critiques Some analysts find the notion of the exchange rate as a weapon to be troubling. Three specific objections can arise: (1) governments' ability to manipulate the exchange rate is limited; (2) the

exchange rate should not be used coercively; and (3) the weapon is rarely used in practice. Consider each of these objections in turn.

First, the objection that governments are not generally effective in managing the exchange rate exogenously applies especially to the active form of the exchange rate weapon. But in the presence of high capital mobility, flexible exchange rates are often driven by herd behavior and expectations, and are thus frequently disconnected from the underlying economic fundamentals. In addition, the foreign-exchange markets often exhibit multiple equilibria more than one exchange rate can be consistent with underlying fundamentals and the market rate is determined by private expectations. When private expectations are easily swayed which is more often the case in the presence of large payments imbalances and conflicts over adjustment governments are more likely to be able to induce a shift from one equilibrium to another. Particularly when the rate moves far from equilibrium, authorities might well coordinate the expectations of private participants by articulating an emerging consensus on the direction of movement.

Government officials can influence financial market expectations about rates and underlying policies, depending on market sentiment, by signaling their desire for a stronger, weaker, or stable currency; forswearing intervention; and intervening. Under some market conditions, such as a profound current account imbalance, an official's "no comment" in the face of a significant exchange rate movement can be interpreted by the market as a clear signal of approval. Conflict over trade policy and market access can enhance the markets' sensitivity to official statements. Thus, even if policymakers have only partial influence over the exchange rate, that influence can be substantial at particular junctures.

The second objection that the exchange rate should not be used coercively springs from the notion of the exchange rate as an economic variable whose movements allow consistency of internal and external balance across national economies. The main response to this objection is that simply ac-

knowledging that the exchange rate is sometimes used coercively does not mean that we endorse its use. The exchange rate weapon is a positive, explanatory concept, rather than reflecting a recommendation or prescription. The normative merit of the instrument depends on the use to which it is put and has varied from one historical case to the next.

As a positive concept, the weapon is an analytical tool of political economic analysis rather than of economic analysis per se. While acknowledging the economic function of the exchange rate, political analysis places the exchange rate in a larger context: an international struggle over the distribution of the costs of balance of payments adjustment. The variable is thus intermediate between the national politics of countries' macroeconomic policy, rather than intermediate simply between their economies. The exchange rate weapon is, in this conception, one of several means by which some states alter the payoffs to policy choices on the part of others.

Finally, in response to the objection that the instrument is rarely used, the exchange rate weapon has been employed at especially critical junctures in modern international economic relations. During the crises of the early 1970s, late 1970s, and mid-1980s, for example, U.S. administrations confronted strong protectionist pressures that were defused in part by the use of the exchange rate to secure adjustment. Had the U.S. Treasury simply let market forces bring adjustment in their own good time and in their own good measure, U.S. trade policy could have slowed the pace of liberalization or even become protectionist. Decisions to deploy or withhold the exchange rate weapon, therefore, can cast a long shadow into the future.

See also balance of payments; beggar-thy-neighbor policies; Bonn Summit; Bretton Woods system; exchange rate regimes; expenditure changing and expenditure switching; foreign exchange intervention; international policy coordination; Louvre Accord; Plaza Accord; Swan diagram

FURTHER READING
On the origin of the concept of the exchange rate weapon and its elaboration, see:

Henning, C. Randall. 1998. "Systemic Conflict and Re gional Monetary Integration: The Case of Europe." *International Organization* 52: 537 73.

. 2005. "The Exchange Rate Weapon and Macro economic Conflict." In *International Monetary Power*, edited by David M. Andrews. Ithaca, NY: Cornell University Press, 117 38.

For historical cases of macroeconomic conflict and cooperation, see:

Cooper, Richard N., Barry Eichengreen, C. Randall Hen ning, Gerald Holtham, and Robert D. Putnam. 1989. *Can Nations Agree? Issues in International Economic Co operation*. Washington, DC: Brookings Institution.

Funabashi, Yoichi. 1988. *Managing the Dollar: From the Plaza to the Louvre*. Washington, DC: Peterson Institute for International Economics.

Helleiner, Eric. 1994. *States and the Reemergence of Global Finance: From Bretton Woods to the 1990s*. Ithaca, NY: Cornell University Press.

Henning, C. Randall. 1994. *Currencies and Politics in the United States, Germany, and Japan*. Washington, DC: Institute for International Economics.

Iida, Keisuke. 1999. *International Monetary Cooperation among the United States, Japan, and Germany*. Boston: Kluwer.

Broad treatments of international monetary power and influence by political scientists include:

Andrews, David M., ed. 2005. *International Monetary Power*. Ithaca, NY: Cornell University Press.

Kirshner, Jonathan. 1995. *Currency and Coercion: The Political Economy of International Monetary Power*. Princeton, NJ: Princeton University Press.

C. RANDALL HENNING

■ exchange rates and foreign direct investment

One of the many influences on foreign direct in vestment (FDI) activity is the behavior of exchange rates. Exchange rates, the domestic currency prices of foreign currencies, matter in terms of both their levels and their volatility. Exchange rates can influence the total amount of FDI that takes place and the allo cation of this investment spending across a range of countries.

Understanding the effects of exchange rates on FDI is important for many reasons. For example, alternative exchange rate arrangements adopted by countries can reduce nominal changes in the rel ative values of currencies, sometimes even merging the currencies completely through a monetary union. For countries trying to attract more FDI, what roles do exchange rate levels and volatility play in determining the attractiveness of these countries for FDI? Can countries make themselves more or less attractive by manipulating the exchange rate?

Exchange Rate Levels When a currency depre ciates, meaning that its value declines relative to the value of another currency, this exchange rate move ment has two potential implications for FDI. First, it reduces that country's wages and production costs relative to those of its foreign counterparts. All else being equal, the country experiencing real currency depreciation has enhanced "locational advantage" or attractiveness as a location for receiving productive capacity investments. By this relative wage channel, the exchange rate depreciation improves the overall rate of return to foreigners contemplating an overseas investment project in this country.

The exchange rate level effects on FDI through this channel rely on a number of basic considerations. First, the exchange rate movement needs to be as sociated with a change in the relative production costs across countries, and thus should not be ac companied by an offsetting increase in the wages and production costs in the destination market. Second, the importance of the relative wage channel may be diminished if the exchange rate movements are an ticipated. Anticipated exchange rate moves may be reflected in a higher cost of financing the investment project, since interest rate parity conditions equalize risk-adjusted expected rates of returns across coun tries. By this argument, stronger effects of exchange rate movements on FDI arise when unanticipated and not reflected in the expected costs of project finance for the FDI.

Some experts dismiss the empirical relevance of the interest-parity caveat. Instead, some argue that there are imperfect capital market considerations, leading the rate of return on investment projects to depend on the structure of capital markets across countries. For example, Froot and Stein (1991) argue that capital markets are imperfect and lenders do not have perfect information about the results of their overseas investments. In this scenario, multinational companies that borrow or raise capital internationally to pay for their overseas projects will need to provide their lenders some extra compensation to cover the relatively high costs of monitoring their investments abroad. Multinationals would prefer to finance these projects out of internal capital if possible, since internal capital is less expensive than borrowed capital.

Consider what occurs when exchange rates move. A depreciation of the destination market currency raises the relative wealth of source country agents and can raise multinational acquisitions of certain destination market assets. To the extent that source country agents hold more of their wealth in own currency-denominated form, a depreciation of the destination currency increases the relative wealth position of source country investors. As the parent company's wealth rises, more financing out of internal capital occurs. The reduced relative cost of capital allows these investors to bid more aggressively for assets abroad. Empirical support for this channel is provided by Klein and Rosengren (1994), who show that the importance of this relative wealth channel exceeded the importance of the relative wage channel in explaining FDI inflows to the United States during the period from 1979 through 1991.

Blonigen (1997) makes a "firm-specific asset" argument to support a role for exchange rates movements in influencing FDI. Suppose that foreign and domestic firms have equal opportunity to purchase firm-specific assets in the domestic market but different opportunities to generate returns on these assets in foreign markets. In this case, currency movements may affect relative valuations of different assets. While domestic and foreign firms pay in the same currency, the firm-specific assets may generate returns in different currencies. The relative level of foreign firm acquisitions of these assets may be affected by exchange rate movements. In the simple stylized example, if a representative foreign firm and domestic firm bid for a foreign target firm with firm-specific assets, real exchange rate depreciations of the foreign currency can plausibly increase domestic acquisitions of these target firms. Again, this channel predicts that foreign currency depreciation will lead to enhanced FDI in the foreign economy. Data on Japanese acquisitions in the United States support the hypothesis that real dollar depreciations make Japanese acquisitions more likely in U.S. industries with firm-specific assets.

Exchange Rate Volatility In addition to levels of exchange rates, volatility of exchange rates also matters for FDI activity. Theoretical arguments for volatility effects are broadly divided into "production flexibility" arguments and "risk aversion" arguments. To understand the production flexibility arguments, consider the implications of having a production structure whereby producers need to commit investment capital to domestic and foreign capacity before they know the exact production costs and exact amounts of goods that will be ordered from them in the future. When exchange rates and demand conditions are realized, the producer commits to actual levels of employment and the location of production. As Aizenman (1992) demonstrated, the extent to which exchange rate variability influences foreign investment hinges on the sunk costs in capacity (i.e., the extent of investment irreversibilities), on the competitive structure of the industry, and overall on the convexity of the profit function in prices. In the production flexibility arguments, the important presumption is that producers can adjust their use of a variable factor following the realization of a stochastic input into profits. Without this variable factor — that is, under a productive structure with fixed instead of variable factors — the potentially desirable effects of price variability on profits are diminished. By the production flexibility arguments, more volatility is associated with more FDI ex ante, and more potential for excess capacity and production shifting ex post, after exchange rates are observed.

An alternative approach linking exchange-rate variability and investment relies on risk aversion arguments. The logic is that investors require compensation for risks, and that exchange rate movements introduce additional risk into the returns on investment. Higher exchange-rate variability lowers the certainty equivalent expected exchange-rate level, as in Cushman (1985, 1988). Since certainty equivalent levels are used in the expected profit functions of firms that make investment decisions today in order to realize profits in future periods, if exchange rates are highly volatile, the expected values of investment projects are reduced, and FDI is reduced accordingly. These two arguments, based on "production flexibility" versus "risk aversion," provide different directional predictions of exchange rate volatility implications for FDI.

The argument that producers engage in international investment diversification in order to achieve ex post production flexibility and higher profits in response to shocks is relevant to the extent that ex post production flexibility is possible within the window of time before the realization of the shocks. This distinction suggests that the production flexibility argument is less likely to pertain to short-term volatility in exchange rates than to realignments over longer intervals.

When considering the existence and form of real effects of exchange rate variability, a clear distinction must be made between short-term exchange rate volatility and longer-term misalignments of exchange rates. For sufficiently short horizons, ex ante commitments to capacity and to related factor costs are a more realistic assumption than introducing a model based on ex post variable factors of production. Hence, risk aversion arguments are more convincing than the production flexibility arguments posed in relation to the effects of short-term exchange rate variability. For variability assessed over longer time horizons, the production flexibility motive provides a more compelling rationale for linking FDI flows to the variability of exchange rates.

As explained earlier, the exchange rate effects on FDI are viewed as exogenous, unanticipated, and independent shocks to economic activity. Of course,

to the extent that exchange rates are best described as a random walk, this view is a reasonable treatment. Otherwise, it is inappropriate to take such an extreme partial equilibrium view of the world. Accounting for the comovements among exchange rates, monetary demand, and productivity realizations of countries is important. As Goldberg and Kolstad (1995) show, these correlations can modify the anticipated effects on expected profits and the full presumption of profits as decreasing in exchange rate variability. Empirically, exchange rate volatility tends to increase the *share* of a country's productive capacity that is located abroad. Analysis of two-way bilateral FDI flows between the United States, Canada, Japan, and the United Kingdom showed that exchange rate volatility tended to stimulate the share of investment activity located on foreign soil. For these countries and the time period explored, exchange rate volatility did not have statistically different effects on investment shares when distinguished from periods where real or monetary shocks dominated exchange rate activity. Real depreciations of the source country currency were associated with reduced investment shares to foreign markets, but these results generally were statistically insignificant.

Although theoretical arguments conclude that the share of total investment located abroad may rise as exchange rate volatility increases, this logic does not imply that exchange rate volatility depresses domestic investment activity. In order to conclude that domestic aggregate investment declines, one must show that the increase in domestic outflows is not offset by a rise in foreign inflows. In the aggregate U.S. economy, exchange rate volatility has not had a large contractionary effect on overall investment (Goldberg 1993).

Overall, the current state of knowledge is that exchange rate volatility can contribute to the internationalization of production activity without depressing economic activity in the home market. The actual movements of exchange rates can also influence FDI through relative wage channels, relative wealth channels, and imperfect capital market arguments.

See also exchange rate regimes; exchange rate volatility; location theory

FURTHER READING

Aizenman, Joshua. 1992. "Exchange Rate Flexibility, Volatility, and Patterns of Domestic and Foreign Direct Investment." *International Monetary Fund Staff Papers* 39 (4): 890 922. Investigates the factors determining the impact of exchange rate regimes on the behavior of domestic investment and FDI, and the correlation between exchange rate volatility and investment.

Blonigen, Bruce A. 1997. "Firm Specific Assets and the Link between Exchange Rates and Foreign Direct Investment." *American Economic Review* 87 (3): 447 965. Uses data on Japanese acquisitions of U.S. industries to show that dollar depreciation raises the likelihood of such acquisitions, especially in industries with high firm specific assets.

Cushman, David O. 1985. "Real Exchange Rate Risk, Expectations, and the Level of Direct Investment." *Review of Economics and Statistics* 67 (2): 297 308. Shows U.S. annual bilateral direct investment to industrialized countries responds significantly to expected dollar depreciation and risk.

———. 1988. "Exchange Rate Uncertainty and Foreign Direct Investment in the United States." *Welt wirtschaftliches Archiv* 124 (2): 322 34. Attempts to clarify the ways that exchange rate uncertainty can affect FDI and tests for these effects on U.S. FDI inflows.

Froot, Kenneth A., and Jeremy C. Stein. 1991. "Exchange Rates and Foreign Direct Investment: An Imperfect Capital Markets Approach." *Quarterly Journal of Economics* 106 (4): 1191 1217. Model showing that when global capital markets have informational imperfections, depreciation lowers the relative wealth of domestic agents and leads to foreign acquisitions.

Goldberg, Linda S. 1993. "Exchange Rates and Investment in United States Industry." *Review of Economics and Statistics* 75 (4): 575 88. Shows how exchange rate level and volatility effects on real investment spending differ over time and across U.S. industries.

Goldberg, Linda S., and Charles D. Kolstad. 1995. "Foreign Direct Investment, Exchange Rate Variability, and Demand Uncertainty." *International Economic Review* 36 (4): 855 73. Model and empirical results show that shares of production capacity located abroad rise as exchange rate volatility rises and becomes more correlated with export demand shocks.

Klein, Michael W., and Eric S. Rosengren. 1994. "The Real Exchange Rate and Foreign Direct Investment in the United States: Relative Wealth vs. Relative Wage Effects." *Journal of International Economics* 36 (3 4): 373 89. Tests the Froot and Stein 1991 model, showing relative wealth effects stronger than relative wage effects of exchange rates on U.S. inward FDI.

LINDA S. GOLDBERG

■ exorbitant privilege

Exorbitant privilege refers to the ability of the United States to finance external deficits by printing its own currency. Countries incur external deficits by spending abroad more than they earn abroad. The conventional measure of this is a country's "current account deficit," which consists of three components: its export earnings minus its spending on imports; its earnings on foreign assets (interest payments on bank deposits and bonds, dividends on stocks, and profits on production abroad minus its payments to foreigners on their holdings of domestic assets), and its receipts of aid and private remittances from abroad minus its payments to foreigners for such aid and remittances.

Beginning in the late 1950s, the United States began running ever-larger external deficits: that is, it was "living beyond its means." By late 2007, the U.S. current account deficit was larger than ever before in its or any other country's history: about $850 billion per year, largely comprised of an excess of imports over exports.

Charles de Gaulle, ever envious of Anglo-American privilege, is widely credited with coining the term, but although he popularized it in speeches, it came, in fact, from Valery Giscard d'Estaing, de Gaulle's lieutenant during the 1960s.

The post World War II arrangements designed by Britain and the United States at the Bretton Woods conference in 1944 placed the U.S. dollar at the core of the international monetary system. Al-

most all noncommunist countries agreed to peg their currencies to the dollar, which was in turn pegged to gold. Not only was the dollar literally as "good as gold"; it was for two decades almost the only currency that was freely convertible into other currencies (Canadian and Australian dollars were exceptions). It was also in goodly supply, given America's near-hegemony in world trade and finance from 1945 until the late 1950s. America produced more than 80 percent of world GDP and exports, and America's trade surpluses were invested in dollar-denominated assets issued by foreigners running trade deficits. Hence U.S. dollars became the world's preeminent choice for both trade and finance.

In short, all countries outside the Soviet sphere that aspired to import or export, or to borrow or lend internationally, had little alternative but to use U.S. dollars. America's exorbitant privilege derived from worldwide demand for a financial asset that could be created at near-zero cost. This led to the complaint that America was able to live beyond its means by simply printing money.

In 2007 almost all of the world's central banks still held at least three quarters of their "foreign exchange reserves" in U.S. dollars, usually in the form of bonds issued by the U.S. Treasury. By the time the Bretton Woods system collapsed in 1971 73, the United States was running a large external trade deficit. This was financed by the trade surpluses of large countries like Germany and Japan. Under Bretton Woods arrangements, these countries had kept their currencies pegged to the U.S. dollar at undervalued exchange rates by directing their central banks to sell deutsche marks and yen, and in the process to buy dollars. Trade surpluses accumulated as foreign exchange reserves in the coffers of their central banks.

After the Bretton Woods system collapsed, Germany, Japan, and other countries allowed their currencies to rise to market values. Nevertheless, global demand for dollar-denominated bonds and other dollar-denominated assets continued to grow. Although flexible-exchange-rate countries did not necessarily accumulate international reserves as rapidly as when they had fixed their rates, most countries returned to fixed- or managed-rate regimes. For example, from 1979 to 1999 (before the introduction of the euro), core Western European countries fixed their rates against one another. International trade grew rapidly, as did central banks' appetite for foreign exchange reserves, mostly in the form of U.S. Treasury bonds. And in the 21st century, China's central bank has become the world's largest single buyer of U.S. Treasury bonds: since 1994, it has held its currency at a below-market exchange rate and has hence run current account surpluses that are mostly invested in U.S. government bonds, enhancing American "exorbitant privilege" as never before.

The United States has imported more than it has exported almost continuously since 1969. America enjoys the exorbitant privilege that it can finance part of this trade deficit simply by printing Treasury bonds, which are readily bought by the world's central banks. To be sure, the United States must pay interest on these bonds, but it is able to pay lower rates than on any other low-risk financial asset the world can offer. Because of the U.S. dollar's near-hegemony in international trade, and also because U.S. Treasury bonds are not only safe but quick and easy to sell (they are both default-risk-free and liquid), central banks and other investors are willing to hold Treasury bonds at rock-bottom interest rates relative to government bonds issued by other countries. This is likely to be true for the foreseeable future since no other country's government bonds are rivals for those of the United States in terms of both their default-free record and their sheer volume and hence liquidity.

Exorbitant privilege does not derive only from central banks' purchases of U.S. Treasuries. Private investors and institutions all over the world hold large parts of their portfolios in U.S.-dollar financial assets: in Treasury bonds, in corporate bonds, and in equities. This is partly because such assets yield good risk-adjusted returns, but it is partly also because the United States is a haven of free-market capitalism: it has never imposed currency controls, and (almost) never frozen funds (Iran's in 1978 79 were a notable exception).

In short, U.S. assets are virtually free from default risk and almost free from political risk. They are not, however, free from currency risk: the U.S. dollar can

and does fluctuate widely against other currencies. Implicit in the term *exorbitant privilege* is that the United States has less incentive than other countries to assuage expectations of currency depreciation because it is assured of a "privileged" market for U.S.-dollar debt and equity, even when such assets seem likely to decline in value against other currencies.

See also Bretton Woods system; convertibility; currency substitution and dollarization; dollar standard; dominant currency; euro; foreign exchange intervention; gold standard, international; reserve currency; special drawing rights; vehicle currency

FURTHER READING

De Long, Brad. 2003. "'Exorbitant Privilege,' or, How Worrisome Is the U.S. Trade Deficit?" *Semi Daily Journal* (August 6). http://econ161.berkeley.edu/movable type/2003 archives/001993.html. De Long's "blog" is a continuing and very readable source of in formation and analysis of the U.S. external deficit.

Economist. 2004. "The Falling Dollar." (November 10). Discusses ongoing concern about U.S. external deficits and the dollar in the context of "exorbitant privilege."

Gourinchas, Pierre Olivier, and Helene Roy. 2005. "From World Banker to World Venture Capitalist: U.S. Ex ternal Adjustment and the Exorbitant Privilege." NBER Conference Paper (August). http://socrates.berkeley.edu/~pog/academic/exorbitant/exorb privi lege 0804.pdf. Links long standing concern about the "exorbitant privilege" of the United States to its more recent role as a "venture capitalist." It also, in passing, gives due credit for the phrase to Giscard d'Estaing.

JAMES W. DEAN

■ **expenditure changing and expenditure switching**

In an open economy setting, policymakers need to achieve two goals of macroeconomic stability: internal and external balances. Internal balance is a state in which the economy is at its potential level of output that is, a country's resources are fully employed and domestic price levels are stable. External balance is attained when a country is running neither an excessive current account deficit nor surplus (net exports are equal or close to zero). Attaining internal and external balances requires two independent policy tools: expenditure changing policy and expenditure switching policy.

Expenditure changing policy, which takes the form of fiscal or monetary policy, aims to affect income and employment with the goal of equating domestic expenditure, or absorption, and production. Expenditure switching is a macroeconomic policy that affects the composition of a country's expenditure on foreign and domestic goods. More specifically, it is a policy to balance a country's current account by altering the composition of expenditures on foreign and domestic goods. Not only does it affect current account balances, but it also can influence total demand, and thereby the equilibrium output level.

Internal and External Balances The interaction between internal and external balances can be demonstrated through a simple Keynesian model where consumption is a function of disposable income; the current account is related to the real exchange rate and disposable income (while foreign income that affects the domestic country's exports is assumed to be constant); and investment and government spending are exogenous. Internal and external balances are:

Internal balance (II):

$$Y = Y^f = C(Y^f - T) + I + G$$
$$+ CA(EP^*/P, Y^f - T)$$

External balance (XX):

$$CA = CA(EP^*/P, Y - T) = XX$$

where XX is a sustainable amount of current account deficit or surplus.

When the exchange rate is flexible, fiscal expansion either government expenditure increase or tax cuts raises output but worsens current account balances. Conversely, a fiscal contraction improves current account balances but lowers output. More specifically, if a country wants to raise its income level through fiscal expansion, it would have to experience a worsening in trade balances, because expansionary

fiscal policy would lead to a rise in imports through improved disposable income and, therefore, worsens current account balances. Alternatively, if a country with a current account deficit attempts to reduce it, it could achieve that by implementing contractionary fiscal or monetary policy, so that as to reduce imports. When a country wants to achieve both internal and external balances simultaneously, it is most effective if the country lets the value of its currency change, so that change in the real exchange rate can affect both the economy's total demand and the demand for imports. Such policy to achieve current account balances by manipulating the demand for domestic and foreign goods through changes in the value of the currency is called expenditure switching policy.

When expenditure switching policy is not available   that is, when an economy is under the fixed exchange rate regime   expenditure changing policy through fiscal policy becomes the only available policy tool for attaining internal and external balances. In the fixed exchange rate system, monetary policy becomes unavailable because it affects the interest rate and the exchange rate. However, fiscal policy is insufficient to achieve both internal and external balances in such an environment.

Effects of Expenditure Changing Policy Although it is expected that expenditure changing policy with fiscal policy changes can affect output in the short run regardless of whether the exchange rate is flexible or fixed, its effect, or the "multiplier of fiscal policy," is smaller in an open economy than in a closed economy. When fiscal expansion is implemented, money demand and thereby the interest rate increase. This rise discourages, or "crowds out," private investment. This outcome arises as long as some degree of price stickiness is assumed. Hence some of the effect of fiscal expansion will be offset by the crowding out of investment, which makes the overall effect on income and also net exports (i.e., $EX - IM = S - I$) smaller than what could have been if the interest rate were assumed to be constant. Also, the multiplier is smaller the more open to international trade the economy is, because a greater portion of income "leaks out" of the system in the form of demand for foreign goods.

Expenditure changing policy with monetary expansion, on the other hand, involves a reduction in the interest rate in the short run, which expands income and worsens net exports. Both types of expenditure-increasing policies function in the same way: income rises while current account worsens in the short run. However, while monetary expansion favors private investment, fiscal expansion favors government spending.

Under the fixed exchange rate system, while monetary policy becomes ineffective, the effect of fiscal policy can be larger than under the flexible exchange rate system. When expansionary fiscal policy is implemented, the interest rate would rise because of the crowd-out effect. At the same time, however, the central bank would have to implement accommodative, that is, expansionary, monetary policy to cancel the rise in the interest rate. The action of canceling the effect on the money supply or interest rate is called sterilization. Otherwise, the interest rate would be affected, and that would affect the capital flows across the border (given the unchanged foreign interest rate) and therefore the exchange rate. Because fiscal expansion must be accompanied with sterilization, the effect of fiscal expansion on output is larger than that under the flexible exchange rate system where the exchange rate is allowed to fluctuate to reflect the change in the interest rate.

Effects of Expenditure Switching Policies Among possible expenditure switching policies, devaluation, or revaluation, is the most focused policy to affect current account balances and the equilibrium level of output. Devaluation increases the domestic price of imports and decreases the foreign price of exports; therefore, it decreases imports and increases exports. However, whether devaluation leads to an improvement in current account balances depends on the elasticities of demand for exports and imports. According to the Marshall-Lerner condition, if the sum of the elasticities of demand for exports and imports is greater than one, depreciation of the domestic currency leads to a current account improvement.

When an economy attempts to attain both internal and external balance, expenditure switching

policy alone is insufficient. For example, if an economy is at the full employment level, in other words, internal balance is already attained, but if it is running current account deficits, policymakers in the economy could devalue its currency so that net exports rise. However, the improvement of current account balances would lead the economy to experience overheating so that internal balance would disappear. If an economy is experiencing an inflationary gap, or overheating, while maintaining a balanced current account, a revaluation policy may reduce total expenditure back to the full employment level, but lead to a current account deficit. Therefore, a mix of expenditure switching and changing policies is usually necessary to achieve both internal and external balances.

With the assumption that the Marshall-Lerner condition holds, for any given level of expenditure, devaluation leads to improvement of net exports, or current accounts, and therefore, a rise in output. However, when prices are assumed to be sticky in the short run, expenditure switching policy with devaluation involves the crowding-out effect. That is, the increase in output also raises the demand for money and consequently the interest rate, which discourages private investment. It is the crowding-out effect that offsets part of the income increase caused by devaluation. Hence, the new equilibrium income level will be a little lower than what could be achieved if the interest rate could remain constant.

Although devaluation policy is the most focused expenditure switching policy, it is not the only one. In general, expenditure policies take the form of trade (control) policy since they are aimed at affecting the volumes of exports and imports. Tariff policy can be implemented to discourage the inflow of imports, and export subsidy can be used to encourage exports, though these policies tend to be industry specific. The best-known tariff policy that has been actually implemented with macroeconomic ramifications is the infamous Smoot-Hawley Tariff Act of 1930. The goal of this policy was to switch demand for foreign goods to domestic ones at the expense of other countries to rescue domestic industries battered by the Great Depression. This policy, however, was followed by other countries that also tried to protect their domestic industries, eventually leading to rapid contraction of international trade.

See also assignment problem; balance of payments; capital mobility; exchange rate regimes; foreign exchange intervention; impossible trinity; Marshall-Lerner condition; money supply; Mundell-Fleming model; New Open Economy Macroeconomics; real exchange rate; sterilization; Swan diagram

FURTHER READING

Caves, Richard E., Jeffrey A. Frankel, and Ronald W. Jones. 2002. *World Trade and Payments An Introduction.* Boston: Addison Wesley. The second half of this book presents theories and models for international macro economics. The effects of government macroeconomic policies are discussed from both the Keynesian and monetarist views as well as for economies with different degrees of capital account openness.

Corden, W. Max. 1969. "The Geometric Representation of Policies to Attain Internal and External Balance." In *Internal Finance*, edited by Richard N. Cooper. Har mondsworth, UK: Penguin Books, 256 90. A seminal paper where the models for expenditure switching and changing policies are introduced.

Krugman, Paul, and Maurice Obstfeld. 2006. *International Economics Trade and Policy.* New York: Pearson/Ad dison Wesley. An introductory textbook for interna tional trade and finance. In the international finance section, simple models for expenditure switching and changing policies are discussed. But for most parts of the book, free flows of capital are assumed.

HIRO ITO

■ export platform foreign direct investment
See vertical versus horizontal foreign direct investment

■ export processing zones
Export processing zones (EPZs) are locations where host countries have established special incentives to

attract producers that import raw and intermediate components, process those items, and then export the final product. The regulatory structure governing EPZ operations in most zones is relatively straightforward. Participating firms are allowed to import inputs and equipment into the host country and export their final product duty free. Host countries may also provide zone firms incentives such as reduced income and property taxes, free or low-cost land, reimbursement for training expenses, and other benefits. In some countries, participating firms have to be located in a specific area. In others, producers qualifying for EPZ benefits can be located anywhere in the country. China, the Philippines, Malaysia, Mexico, Costa Rica, El Salvador, Honduras, Tunisia, Mauritius, and Madagascar are just a few of the many developing countries that have been able to dramatically increase exports through their EPZ initiatives. As a result, EPZs have become increasingly popular in the developing world. The International Labor Organization (ILO) reports that in 1997, 93 countries had established zones and total EPZ employment stood at 22.5 million. By 2002 these numbers had grown to 116 countries and 43 million jobs.

EPZs and Host Country Development EPZs are a controversial aspect of the global economy. On the positive side, they offer clear benefits to both producers and host countries. For producers, EPZs provide multinational companies (MNCs) and others easy access to a low-cost workforce in a developing country. For host countries, EPZs help create jobs, generate foreign exchange, and increase exports. EPZ opponents, however, contend that economic development strategies based on EPZ initiatives result in host countries voluntarily assuming an unattractive position within the global production system. MNCs have historically kept high-value-added, high-complexity links in the value chain close to the corporate center and delegated fragmented, low-value-added assembly tasks to their developing country operations. EPZ firms tend to be concentrated in labor-intensive industries such as apparel, footwear, toys, and electronics assembly, and to employ 18- to 25-year-old women with little education or formal training.

In developing countries where host country governments are struggling to provide the basic necessities of life to their population, low-value-added EPZ activity may represent a positive contribution to national development. To constitute a positive force in more advanced developing countries, the EPZ sector must transition toward a higher-value-added, higher-complexity production model. There are clear examples of countries where a significant percentage of EPZ firms have made this transition. This evolution can also include the transfer of knowledge-intensive support activities such as research and development and design engineering to the host country location. In other zones, the majority of firms have not evolved beyond basic assembly, and wages, especially for operating-level workers, have stagnated.

The quality of the jobs created through EPZ initiatives has been a consistent point of contention between EPZ proponents and their critics. The ILO, nongovernmental organizations, and academic researchers all express concerns about and actively monitor employment conditions in EPZs. In most countries it appears that zone firms provide at least equal if not better terms of employment than local companies, especially when compared to small businesses and the informal sector. There is, however, a clear need to develop appropriate regulatory structures and to aggressively monitor and enforce EPZ labor and environment standards. Of particular concern for the ILO, many developing countries restrict the opportunity for EPZ employees to join representative unions and engage in collective bargaining.

Critics also argue that the business model pursued by EPZ firms creates export enclaves with few if any linkages to other sectors of the host country economy. A closer look at day-to-day operations in successful zones shows that EPZs create numerous opportunities for construction companies, transportation firms, customs brokers, and a host of additional host country service providers. Local entrepreneurs also have the opportunity to supply raw and intermediate inputs that zone producers would otherwise import from global sources. For a variety of

reasons, however, local firms have been unable to take full advantage of these opportunities in many developing regions. In Mexico, for example, inputs from national suppliers represent less than 4 percent of all raw and intermediate components used by the EPZ sector.

EPZs and Their Impact on the Developed World The use of EPZs by producers in many industry segments creates clear benefits for consumers in the developed world. Through accessing an abundant, low-cost workforce, companies are able to sell products at relatively low prices while still generating a profit. A number of well-known retailers such as Wal-Mart and Home Depot have increasingly turned to sourcing goods from EPZ plants in China as a core component of their business model. China's emergence as the world's premier EPZ location is frequently cited as one of the main reasons inflation has remained low in the developed world over the last decade.

As with so many other aspects of globalization, these benefits come at a cost. The argument has been made that EPZs represent one of the primary forces resulting in a so-called race to the bottom in global labor and environmental standards. Critics typically support this contention with three main points. First, EPZs clearly facilitate the exodus of manufacturing jobs from high-wage developed countries to low-wage developing regions. Second, the EPZ option also reduces the power of organized labor to bargain for improved wages and working conditions in industries where it is possible for companies to shift production to low-wage EPZs. By contributing to reduced demand for workers and the disruption in the bargaining power between management and unions, EPZs may be partially responsible for the recent stagnation in blue-collar wages in the United States and elsewhere. Finally, critics also contend that MNCs frequently move environmentally risky operations to developing country EPZs where appropriate standards are lacking and/or enforcement is weak. There is considerable, primarily qualitative evidence for the first two points, but little if any sound support indicating that large numbers of MNCs have moved to developing country EPZs to avoid strict environmental standards.

China's New Role in the Global EPZ Industry The primary location of EPZ-style production was very different in the early 2000s compared to prior decades. Starting from a very low base in the late 1970s China has rapidly become the world's preferred EPZ location. In 1997, the ILO estimated 18 million people were employed in Chinese EPZs. In December 2001, China gained full membership in the World Trade Organization (WTO) and with improved market access to the world's primary industrial and consumer markets EPZ employment increased to 30 million by 2002. This total represents close to 70 percent of the world's EPZ workforce.

EPZ producers are the main force responsible for China's increased prominence in world markets. It is estimated that EPZ firms account for more than half of all Chinese exports and 70 percent of Chinese exports to the United States. China's attraction as a production location is due to several factors: the undervalued Chinese currency, the availability of large numbers of very low cost workers as well as skilled middle- and upper-level engineers and managers, generous incentives provided by the Chinese government, considerable investment in export infrastructure, and the opportunity for firms to simultaneously engage in EPZ-style production and sell a percentage of factory output to the country's 1.2 billion consumers.

China's success is forcing countries in both industrialized and developing regions to rethink their export promotion strategies. A study by Lall and Albaladejo (2004) found higher-cost Asian countries were losing market share to China in the 1990s primarily in low-technology segments and were increasingly focusing on the production of intermediate components that were then exported to China for final assembly. Lall characterizes the trade relationship between China and its East Asian neighbors as complementary rather than competitive. This may change as China upgrades toward more technology-intensive, high-complexity sectors.

China does represent a direct threat for many EPZ producers in Mexico and the Caribbean Basin. The brief recession in the United States and increased competition from Chinese EPZ exports contributed to the loss of close to 300,000 jobs in *maquiladoras*, as Mexican EPZs are known, from 2000 to 2003. Mexican policymakers have argued that China's low-cost advantages will force Mexico's EPZ industry to shift from labor-intensive to more capital- and technology-intensive activities. The available evidence indicates, however, that *maquiladoras* have responded by adopting business models in which competitive advantage is based on proximity to industrial and consumer markets in the United States, such as mass customization, just-in-time delivery, low volume high mix production systems, and remanufacturing centers.

Policy Implications The historical record clearly shows both positive and negative outcomes associated with EPZs. Using scarce development resources to establish a zone in which participating firms employ a very young, unskilled, primarily female workforce at subsistence wages to perform low-value-added assembly work may be worthwhile only in the poorest of countries. This type of participation in the global economy is useful only if it represents a starting point leading to something better. Governments in Asia, Mexico, and elsewhere have employed EPZs as an integral component of much larger and more comprehensive upgrading strategies. Global competition as well as the evolution of the organizational capacity of MNCs present interesting new challenges for host countries with EPZ programs. In search of efficiency and profitability, MNCs continue to migrate entire industries from high- to low-wage locations. Companies are also improving their ability to divide value chains into smaller and smaller segments and place activities around the world where they can be performed most efficiently. The world's largest MNCs are also much more willing than they were in the past to delegate knowledge-intensive functions such as design engineering to their developing country subsidiaries. The offshore outsourcing of service functions is a closely related phenomenon that represents significant new opportunities for developing countries. In Mexico there are now locations where traditional EPZ activity, engineering centers, and call centers are all located in the same industrial park. Based on both the historical record and new opportunities, EPZs have a place in the policy kit of those in developing countries attempting to create wealth through open engagement with the global economy.

See also export promotion; fragmentation; International Labor Organization; outsourcing/offshoring

FURTHER READING

Cling, Jean Pierre, Mireillie Razafindrakoto, and Francois Boubaud. 2005. "Export Processing Zones in Mada gascar: A Success Story under Threat?" *World Develop ment* 33 (5): 785 803. A firsthand account of EPZ ac tivity in Madagascar since 1990.

Grunwald, Joseph, and Kenneth Flamm. 1985. *The Global Factory: Foreign Assembly in International Trade*. Wash ington, DC: Brookings Institution Press. One of the first relatively comprehensive studies of EPZs.

International Labor Organization. 2003. *Employment and Social Policy in Respect of Export Processing Zones* (*EPZs*). BG.286/ESP/3. Downloadable from www.ilo.org. A summary of EPZ activity around the world and ILO labor concerns.

Kaplinsky, Raphael. 1993. "Export Processing Zones in the Dominican Republic: Transforming Manufacturers into Commodities." *World Development* 21 (11): 1851 65. A well thought out critique by a prominent devel opment scholar.

Lall, Sanjaya, and Manuel Albaladejo. 2004. "China's Competitive Performance: A Threat to East Asian Manufactured Exports?" *World Development* 32 (9): 1441 66. An empirical study of the impact of China's exports on the exports of its neighbors.

Lemoine, Françoise, and Deniz Ünal Kesenci. 2004. "As sembly Trade and Technology Transfer: The Case of China." *World Development* 32 (5): 829 50. An excel lent study examining EPZ activity in China.

Romero, Ana Teresa. 1998. "Export Processing Zones in Africa: Implications for Labour." *Competition and*

Change 2 (4): 391 418. An interesting review of EPZs in Africa.

Steinfeld, Edward. 2004. "China's Shallow Integration: Networked Production and New Challenges for Late Industrialization." *World Development* 32 (11): 451 96. Examines the development challenges for China as a result of the emergence of modular production in EPZs.

JOHN SARGENT

■ export promotion

Governments often seek to promote exports from their country and to encourage national production of goods that might be exported. Export promotion can be viewed, therefore, as either an aim of economic policy (an export promotion strategy) or the set of policies and measures used by government agencies and the private sector to promote exporting and the production of exportable goods and services (export promotion policies).

Trade economists and policymakers have long been interested in how the trade policies, institutions, and trade strategy a country adopts affect its export and wider economic performance. The issues that generate this interest are controversial. Differences of opinion exist on the relationship between trade in general, and exports in particular, and economic growth; the influence of trade strategy in general, and an export-promoting strategy in particular, on export performance; and whether and how governments should interfere with market processes in general and in trade sectors in particular.

Traditional neoclassical theory represents trade as exposing domestic markets to greater competition, which induces efficient use and allocation of resources in line with relative world prices. In the absence of pervasive market failures, this representation provides a rationale for limiting government intervention and for adopting relatively neutral government interventions in the traded goods sectors. This resulting openness of the economy to international trade and competition implies in turn an outward orientation of the country's trade strategy, a will-ingness for as much specialization of domestic production and as much growth of imports and exports to occur as is consistent with the openness of the economy.

It is necessary, however, to recognize that the issues of trade orientation and degree of government intervention in markets are analytically separate. Outward orientation and export promotion can be led by markets or by policy intervention. An inward orientation can be removed by dismantling protectionist, import-substituting trade policies and moving to a free trade regime. The removal of import tariffs and other nontariff barriers can be seen as a market-led approach to export promotion. Alternatively, greater incentives to produce for export markets can be offered through additional government measures to subsidize exports and production in export sectors explicitly or implicitly. The relative merits of these alternative approaches to export promotion have been at the heart of the academic and policy debate in this area of international economics, in particular in the context of developing countries' policy options. Given the tendency of governments to use a mixture of import liberalization and direct export promotion measures to implement more outward-oriented trade strategies, controversy has developed over how to interpret the empirical evidence on the relationships among trade strategy, export performance, and economic growth or development.

The Case against Export Promotion There is a strong strand of the development economics literature that rejects the neoclassical focus on domestic supply conditions as an influence on export performance (see Evans 1990). This strand views the level of external demand as critical in influencing the export performance of developing countries. Pessimism about the growth of demand reflects the relatively low income elasticity of demand for primary products and the belief that the growth of manufactured imports by these countries would be met by protectionist responses in the industrial countries. Export pessimism provided a rationale for reducing reliance on external demand and export conditions

for growth and development and led in the post World War II period to the widespread adoption of interventionist and inward-looking policies.

Import restrictions were adopted to protect industries producing for the domestic market, with the aim of causing the demand for home goods to grow by replacing or substituting for imports. An import-substitution trade strategy was implemented through a range of direct and indirect means, which subsidized the production of importable goods, taxed the production of exportables, or did both. Taxes on imports, such as tariffs, or quantitative restrictions on importing (e.g., import licenses or quotas) implicitly subsidized importables production, while export taxes explicitly taxed exportables. Equivalent effects could be achieved also by less direct means. Export marketing boards might use their monopoly position to give commodity exporters less than the world price for their exports. Reluctance to alter a pegged exchange rate in the face of high inflation may cause currency overvaluation, reducing the price competitiveness of exports in foreign markets.

The Case for Export Promotion There are a priori arguments for encouraging outward orientation based on the benefits of more efficient allocation and use of resources due to greater competition. These may be one-time gains that raise per capita incomes but do not alter the steady-state or long-term growth rates of economies. More recently, theoretical support for export promotion has been provided by the so-called new or endogenous growth theories. In these, export growth may, in the presence of increasing returns and complementarities between physical and human capital accumulation, lead to permanent, positive growth effects.

It was not the advent of this new branch of theory, however, that caused most economists and policymakers to reject import-substituting trade strategies. In part they responded pragmatically to the revealed effects of import substitution and the evidence from developing countries of the resulting high costs of protection. Under import substitution, factories were underused, production technologies were often overly capital intensive, and the growth of

trade regulations created high returns to lobbying and other forms of rent-seeking. The move away from import substitution was also in part a reaction to the evidence that some developing countries, particularly in Southeast Asia, had developed an export capacity in manufactured goods and had brought about rapid growth and industrialization (Trindade 2005). This undermined the generality at least of the "export pessimism" arguments for inward orientation (Bhagwati 1990).

Finally, even where policymakers remained unconvinced by the relative performance of inward- and export-oriented countries or by the benefits of moving from one strategy to the other, the macroeconomic and debt problems of their economies forced them into accepting trade policy reform as part of the conditionality for the loans they needed from the World Bank. Structural adjustment programs that were required in order to access these loans were a major vehicle during the 1980s and 1990s in driving the import liberalization (reduction and/or elimination of nontariff barriers, and rationalization and lowering of tariff rates) and exchange rate adjustment (currency devaluation or depreciation). These reforms significantly lowered the large anti-export bias that had grown up in the trade regimes of many developing countries.

Export Opportunities and Mechanisms for Export Development Globalization and the growth of the global economy have brought increased export market opportunities. In addition, the successful newly industrializing and exporting countries demonstrate a "ladder of development" effect. There is evidence of the richer and more industrialized countries being displaced as exporters first of standardized labor-intensive manufactures and subsequently of less standardized and more capital-intensive products. This displacement has been by developing countries of varying size. The emergence of new, large-scale exporters such as China, India, and Mexico may cause other smaller, potential late entrants to worry that there are forces of agglomeration of international production that will make future entry into exporting more difficult. These new

exporting countries, however, also offer additional market opportunities for final and intermediate goods. Furthermore, the increasing fragmentation of international production increases these market opportunities in both developed and developing country markets. Indeed, the key focus of the literature on export development has shifted away from concern about the absence of export opportunities when all countries try simultaneously to promote exports (often referred to as the "fallacy of composition" problem) to identifying the internal and external constraints on export development.

In the 1950s and 1960s attention tended to focus on external constraints, including on the issue of market access to the industrial countries. High trade barriers, especially against labor-intensive imports such as textiles, clothing, and footwear, were seen as constraining export growth in the developing countries. This led to the incorporation of "special and differential" treatment principles for developing countries in the General Agreement on Tariffs and Trade, which allowed the industrialized countries to offer preferential access to their markets for developing country exports. The benefits of preferences (lower tariffs than those applied to exports from non-preference-receiving countries) have been weakened subsequently by the multilateral trade liberalization that the industrialized countries have implemented. They have lowered their tariff and nontariff barriers significantly in general, though market access can still be highly constrained in specific products where contingent protection measures (e.g., anti-dumping controls) and regulations (e.g., technical and health standards) restrict imports.

The continuing small share of world trade of the developing countries led to the shift of attention after the 1970s to the role of the developing countries' own trade policies in constraining export development. As outlined earlier, high import barriers and other protectionist measures created major sources of antiexport bias. Unilateral and regional trade reforms in recent decades have significantly reduced this constraint on export development. Before they liberalized their imports, however, many developing countries experimented with trying to simultaneously import-substitute and export promote, that is, to supplement existing import-substitution measures with fiscal incentives for exporting.

In broad terms simultaneous promotion of import-substituting and exportables production may well be mutually offsetting, moving the economy toward similar relative prices as under free trade if the incentives to both sectors are of similar magnitude. Taxing imports and subsiding exports will, however, absorb resources to collect and spend the taxes. The costs of policy administration can be avoided with free trade. Nonetheless, governments were often reluctant to liberalize imports for ideological and political economy reasons.

One strategy pursued by reforming governments was to continue with a general import substitution regime, while creating enclaves of production for export markets. Domestic and potentially footloose foreign firms within these enclaves, or export processing zones, often gained access to duty-free imported inputs and capital goods or to subsidized inputs and were subject to preferential profits tax arrangements not applied to firms in the rest of the economy. Such arrangements tend to give rise to monitoring problems associated with trying to avoid leakage of subsidized production into the domestic market (Warr 1990). It is also increasingly recognized that exporting requires access to reliable and low-cost inputs and services, which is more likely where exporters are integrated into competitive, open economies.

Recently, attention has shifted from trade policy constraints to other policy and nonpolicy constraints on export development. Increasingly advocates of outward orientation of trade policies recognize that trade policy reform may not be a sufficient condition for improved export performance. Export promotion may well require improved institutions (to support better export marketing or provide cheaper trade credits) and improved infrastructure (better roads for internal sourcing of inputs and distribution or cheaper and more reliable international transportation services). There is growing evidence that

the "taxing" effects on exports that arise from developing countries' own trade policies or from the trade policies of their trading partners are generally less important than those resulting from the inefficiencies or inadequacies of local services and input suppliers, of local institutions in both the private and public sectors, and of the national and regional communications and transport infrastructures (see Milner 2004). Export promotion is in fact recognized as being a much broader concept than that of the country's trade policy stance alone. It encompasses measures to reduce production costs in general and to reduce transportation (internal and international) costs in particular.

See also export processing zones

FURTHER READING

Bhagwati, Jagdish N. 1990. "Export promoting Trade Strategy: Issues and Evidence." In *Export Promotion Strategies: Theory and Evidence from Developing Countries*, edited by Chris Milner. New York: New York University Press, 11 39. A clear statement of this proponent's case for export promotion.

Evans, H. David. 1990. "Outward Orientation: An Assessment." In *Export Promotion Strategies: Theory and Evidence from Developing Countries*, edited by Chris Milner. New York: New York University Press, 40 58. A critical assessment of the case for export promotion, with support for the case against.

Milner, Chris R., ed. 1990. *Export Promotion Strategies: Theory and Evidence from Developing Countries*. New York: New York University Press. A collection of essays on the theory and evidence relating to export promotion, with a section on alternative instruments of export promotion and on national case studies.

———. 2004. "Constraints to Export Development in the Developing Countries." In *The WTO and Developing Countries*, edited by Homi Katrak and Roger Strange. London: Palgrave, 213 32. Provides a useful broad evaluation of the relative importance of market access, trade policies, and other nontrade policy barriers in constraining export development.

Trindade, V. 2005. "The Big Push, Industrialization, and International Trade." *Journal of Development Economics* 78: 22 48. Provides a theoretical analysis of the industrialization of countries such as South Korea and Taiwan produced by export promotion policies.

Warr, P. 1990. "Export Processing Zones." In *Export Promotion Strategies: Theory and Evidence from Developing Countries*, edited by Chris Milner. New York: New York University Press, 130 61. A useful paper on the costs and benefits of export processing zones.

CHRIS MILNER

■ expropriation risk

See trade costs and foreign direct investment

■ factor endowments and
foreign direct investment

International factor endowment differentials (FEDs) between home and host countries of multinational corporations (MNCs) setting up affiliate subsidiaries are one of the key determinants of foreign direct investment (FDI) flows. One reason that some corporate operations are geographically separated from headquarters activities is to minimize production costs. Aspects of production that are intensive in a particular factor (such as labor or a natural resource input) tend to be located where that factor is abundant.

Ascertaining how FEDs affect FDI is important for achieving a general understanding of the determinants of FDI. There is a broad concern, in developed countries, that production is fleeing high-cost industrialized countries for low-cost developing countries such as China that FDI is motivated by firms trying to slash costs by shifting labor-intensive activities to countries where labor is cheap.

Yet it is important to recognize that to operate in a given host country, in addition to requiring abundant labor to make production costs low, the subsidiary of an MNC may also need to import machinery and employ engineers. Hence while relative labor abundance makes wage production costs low, the other production costs, such as for physical capital and human capital, cannot be exceedingly high for the country to be an attractive host for affiliate production. Moreover, the host location for affiliate operations must have nonproduction costs that also are sufficiently low. In fact, the poorest

countries attract almost no FDI. Considering all costs of bringing a good to market, not just labor production costs, poorer countries are often not cheaper locations.

This entry will consider theories of how FEDs affect FDI, and the empirical evidence. Understanding this issue will help form the basis for a better overall knowledge of why FDI occurs.

Global Firm Organization and Modes of FDI FDI occurs when a firm decides to locate production abroad and opts to maintain organizational, as well as financial, control over its foreign production. The first step in the choice to undertake FDI is the decision to relocate part of the production process offshore. The second step relates to organization whether the foreign production is to be organized within the confines of the corporation or at arm's length. In the former case, the corporation is multinational and sets up an affiliate subsidiary outside the country where its headquarters are located. If the costs of setting up the affiliate subsidiary are financed via the MNC's headquarters, there is a financial inflow into the host country in the form of FDI. When the overseas production is at an arm's length operation rather than a subsidiary, there is foreign outsourcing rather than FDI.

In pursuing FDI, MNCs deploy various modes of entry, with the traditional ones being vertical and horizontal. The mode of entry will impinge on the role of FEDs in determining FDI, as each mode of entry seeks to minimize a particular cost to the MNC. First, *vertical* affiliate entry breaking up stages across countries seeks manufacturing cost

minimization through production line fragmentation. Second, *horizontal* subsidiary entry creating duplicate plants abroad aims for transportation cost minimization by bringing production closer to customers. Third, using a host country as an *export platform* by a MNC building a plant abroad to serve a third market balances advantages emanating from FEDs enhanced by geographic location advantages minimizing trade impediments, including tariffs.

Vertical FDI involves different locations for the various parts of the production process of the MNC's final output. This type of organization involves international trade between the firm's subsidiaries with headquarters, and seeks to exploit international factor price differentials. As an example, a MNC might locate headquarters in a skilled labor abundant home country and engage in unskilled labor intensive production in an unskilled labor abundant host country. In this case, there are foreign affiliate exports of final output from the host to the home country of the MNC. Vertical FDI is expected to be more prevalent the greater the FEDs across countries. Greater FEDs give rise to greater differences in factor prices and thus greater incentive for vertical FDI.

Horizontal FDI minimizes the costs of serving the market by producing locally rather than trading. It is more likely when trade costs are relatively high (e.g., tariff-jumping FDI). Establishing a plant in the host country incurs additional plant-level fixed costs but avoids per unit trade costs (including both tariffs and transportation costs), and hence is preferable when the host country market size is sufficiently large. Using a monopolistic competition model, Markusen and Venables (2000) demonstrate that multinationals are more likely to arise when countries are more similar in both relative and absolute endowments. At the same time, a reduction in transportation costs for intermediates can lead to either vertical FDI (when upstream activities are labor intensive) or horizontal FDI (when downstream activities are labor intensive).

The vertical and horizontal modes of FDI have been combined in the "knowledge-capital" model of the multinational firm by Carr, Markusen, and Maskus (2001). Barriers to investment and trade between countries give rise to endogenous FDI modes. The higher the relative costs imposed by trade barriers relative to investment barriers, the more likely MNCs are to engage in horizontal rather than vertical FDI. At the same time, MNCs need not view horizontal and vertical FDI as mutually exclusive strategies. In some cases, a host FDI country may play the role of export platform to serve a regional market.

When there is possible geographic fragmentation in the stages of production into intermediate input manufacturing and assembly, interesting complementarities emerge that give rise to more complex organizational forms.

First, Yeaple (2003) identifies one complementarity that results when the lower cost attainment associated with production relocation induces a rise in overall output, which in turn will give rise to further production relocation. Moreover, Grossman, Helpman, and Szeidl (2003) show that when transportation costs for final goods are moderate, there is a complementarity associated with the source of components: the cost savings from locating assembly in the low-cost region are higher when components are also produced there. This complementarity implies that FDI in assembly responds to changes in the costs of FDI in components. When transportation costs for components are high enough, there is also an agglomeration complementarity in FDI: firms conducting assembly abroad want to locate component production nearby to reduce the costs of shipping components. In addition, the interest in hybrid integration strategies has led to models with more than two countries, with emphasis not only on FEDs but also trade and investment costs across regions of the world.

When producing overseas raises MNC profitability, FDI rather than offshoring will be more likely, the more contractual difficulties the multinational corporation faces in the country where component production is most efficient. Hence the decision to engage in FDI jointly minimizes production and transaction costs for the investor. In this context, Antras and Helpman (2004) show that,

other things equal, countries with a level of contractual frictions will receive more FDI inflows, the more abundant their endowment of the factor used intensively in production of intermediate inputs.

In terms of the financing of production once the offshoring decision is made, Antras, Desai, and Foley (2007) develop a model and provide evidence suggesting that the less developed investor protection is in a potential host country of FDI, the more of an outside stake by the MNC will be required in local capital markets to fund the affiliate subsidiary. More funding for the affiliate in the form of FDI will be forthcoming from headquarters, the less investor protection and financial development there is in the host country.

Thus given that relative factor abundance makes a country comparatively attractive to locate some stage of production activity, the inflow of FDI into that country should, controlling for other relevant country characteristics, increase with higher contractual frictions and lower investor protection. To some extent, FDI is a way to organize production within the confines of the MNC to substitute for poor institutions and incomplete markets in the host country where manufacturing can be most productive due to FEDs, or other motives for production relocation.

Evidence on FEDs and FDI Based on simple models of MNC behavior, FEDs should give rise to more vertical, but less horizontal, FDI. Brainard (1997) reports findings that are inconsistent with multinational activity depending on factor proportion differences. The finding is that increases in per worker income differences reduce affiliate sales (absolutely and relative to trade). Estimates of the knowledge-capital model of the multinational enterprise, which embeds both horizontal and vertical FDI, find support for horizontal FDI.

However, several recent studies do lend support to vertical FDI models based on FEDs. Hanson, Mataloni, and Slaughter (2005) find that demand for imported inputs increases when affiliates face lower wages for less-skilled labor. Braconier, Norback, and Urban (2005) find that more FDI is conducted in countries where less-skilled labor is relatively cheap.

Given the recent reductions in transportation costs, facilitating both international trade of goods and factor flows, traditional theories would predict that FEDs' importance in determining FDI would be mitigated. To the extent that FEDs play an important role in explaining recent FDI flows, this is likely to be due to the interaction, emphasized in more recent theories, of FEDs with some other determinant of FDI. For example, the complementarities discussed previously predict that a fall in transportation costs within an intermediate range may exacerbate the impact of FEDs on FDI flows.

More generally, theories on complex modes of FDI suggest that bilateral sales by MNCs between two countries are not determined solely by the characteristics of those two countries but are also affected by characteristics of other countries and regions. There is evidence consistent with bilateral FDI being complementary with exogenous bilateral and third-country determinants in the case of horizontal export-platform FDI. For vertical FDI, bilateral and third-country effects of changes in skilled and unskilled labor endowments tend to be substitutes for complex vertical FDI. Indeed, more recent evidence suggests that international FEDs matter for FDI not only at the bilateral level but in a more general sense.

Much of the recent expansion in trade flows would appear to be related to the growth of FDI. MNCs account for approximately half of U.S. exports. These are intrafirm transactions between the headquarters of U.S. MNCs and their foreign affiliates (Hanson, Mataloni, and Slaughter 2005). Helpman, Melitz, and Yeaple (2004) have shown that the effect of sector- and country-specific trade impediments in favoring FDI over exports is more pronounced under firm heterogeneity, both in theory and data. In addition, there is evidence that the most productive MNCs are able to atomize production facilities across countries and are not bound to central locations. In global markets, the organizational changes due to firm growth have overall been associated with higher FDI in response to international FEDs.

Taken together, these facts imply that the growth in tandem of FDI and exports reveals the increasing

global importance of MNC activities and their fragmentation across countries. The most productive MNCs have far reach across world markets both in intermediate input production organization and in assembly for global supply of final goods. The cost advantages conferred by certain locations due to FEDs are magnified for the select MNCs whose FDI flows, in their many modes, have global presence.

See also location theory; multinational enterprises

FURTHER READING

Antras, Pol, Mihir Desai, and Fritz Foley. 2007. "Multi national Firms, FDI Flows, and Imperfect Capital Markets." NBER Working Paper No. 12855. Cam bridge, MA: National Bureau of Economic Research. Models and provides evidence on the host country de terminants of how MNCs finance offshore operations.

Antras, Pol, and Elhanan Helpman. 2004. "Global Sourc ing." *Journal of Political Economy* 112 (3): 552 80. Develops a theoretical framework on the contractual environment, transaction cost, and production cost differences across countries in determining the organi zation of firms in a global economy.

Braconier, Henrik, Pehr Johan Norback, and Dieter Ur ban. 2005. "Multinational Enterprises and Wage Costs: Vertical FDI Revisited." *Journal of International Eco nomics* 67 (2): 446 70. Provides empirical evidence about the impact on FDI of the wage differential be tween MNC home and host countries.

Brainard, S. Lael. 1997. "An Empirical Assessment of the Proximity Concentration Trade off between Multi national Sales and Trade." *American Economic Review* 87 (4): 520 44. Analyzes empirically the link between FDI and MNC exports from the home country to the host country.

Carr, David L., James R. Markusen, and Keith E. Maskus. 2001. "Estimating the Knowledge Capital Model of the Multinational Enterprise." *American Economic Review* 91 (3): 693 708. Lays out a framework in which MNCs choose the mode of FDI and provides empirical evi dence.

Grossman, Gene M., Elhanan Helpman, and Adam Szeidl. 2006. "Optimal Integration Strategies for the Multi national Firm." *Journal of International Economics* 70 (1): 216 38. Models how the possibility of intermediate

input trade enriches the choices of MNC operation location including exports to third countries from host countries for FDI.

Hanson, Gordon, Raymond Mataloni, and Matthew Slaughter. 2005. "Vertical Production Networks in Multinational Firms." *Review of Economics and Statistics* 87 (4): 664 78. Provides an econometric analysis of the pattern of international trade and investment within U.S. MNCs.

Helpman, Elhanan, Marc Melitz, and Stephen Yeaple. 2004. "Exports versus FDI with Heterogeneous Firms." *American Economic Review* 94 (1): 300 16. Models and explores empirically how heterogeneous productivity across firms affects the trade off for firms between export and FDI strategies.

Markusen, James R., and Anthony J. Venables. 2000. "The Theory of Endowment, Intraindustry and Multi national Trade." *Journal of International Economics* 52 (2): 209 34. Provides a model to assess the link between FEDs and the role of MNCs in international trade.

Yeaple, Stephen. 2003. "The Complex Integration Strate gies of Multinationals and Cross Country Dependencies in the Structure of Foreign Direct Investment." *Journal of International Economics* 60 (2): 293 314. Provides a model in which vertical and horizontal FDI are not mutually exclusive and do not exhaust the range of market penetration strategies.

MAURICE KUGLER

■ **factor proportions**
See Heckscher-Ohlin model

■ **fair trade**

The concept of fair trade carries an implicit critique of the *unfairness* associated with the conventional North-South trade system and the proposition that an alternative *fairer* trade system is possible. Fair trade advocates reject the assumption that an un regulated, comparative advantage based trade sys tem provides a viable avenue for sustainable devel opment with most artisans, small-scale farmers, and

workers. The persistent poverty among small-scale agricultural commodity producers and artisans in the global South demonstrates the need for a different type of trading system. Since the 1940s, fair trade organizers have worked to create and expand an alternative trade system that starts with a set of commonly held principles codified into standards intended to support empowerment, gender equity, long-term partnerships, transparency, and sustainable community development (FLO 2007; Barratt-Brown 1993). According to an alliance among several international fair trade associations, "Fair trade is a trading partnership, based on dialogue, transparency and respect, which seeks greater equity in international trade. It contributes to sustainable development. Fair trade organizations (backed by consumers) are engaged actively in supporting producers, awareness raising and in campaigning for changes in the rules and practice of conventional international trade" (Krier 2005).

In practice, fair trade is an uneven global movement and an expanding market. The twin strategies for implementing fair trade principles are the creation of a market that offers better prices to small-scale producer organizations and support from international development nongovernmental organizations (NGOs) for business and social development. Fair trade organizations have established an international labeling system (Fair Trade Labeling Organizations International) and international associations of alternative trade organizations (IFAT) to implement these practices, expand awareness among in Northern countries, and build demand.

Although fair trade accounts for 1 5 percent of the global trade in specific agricultural commodities and handicrafts, the markets have expanded rapidly. As of October 2006, the global fair trade certified network included 586 producer organizations and 1.4 million farmers, artisans, and workers in 58 developing countries from Latin America, Asia, and Africa (FLO 2007). Consumers worldwide spent U.S. $2.024 billion on fair trade certified products (FLO 2007). The expanding list of fair trade products includes coffee, cocoa, tea, fruits, wine, sugar, honey, bananas, rice, crafts, and some textiles (Krier 2005). Although artisans and craft producers helped to pioneer the fair trade movement, coffee has emerged as the economic backbone of the certified fair trade system (Bacon et al. 2008).

Fair Trade History The historical roots of the global fair trade movement, like those of many international social movements, extend into many lives, organizations, and landscapes. The primary partnership consisted of impoverished small-scale producers and artisans selling high-quality goods to Northern volunteers, NGOs, and, later, businesses. The risk was often shared between producing and consuming ends of the movement: producers and artisans sometimes provided their products months or even years before receiving full payment after volunteers and alternative trade organizations sold them to distant markets. On the other hand, fair trade organizers from the North provided producers with loans that would otherwise be unavailable and bought crafts and coffee before they had developed markets.

The early alternative trade organizations emerged around handicrafts, often connecting religious and politically motivated Northern groups with groups of women artisans in impoverished communities. In 1946, Edna Ruth Byler, a church volunteer working with the Mennonite Central Committee in Pennsylvania, started buying quilts directly from seamstresses in Puerto Rico. Several years later, the church assumed this project and later created Ten Thousand Villages, an alternative trade organization that as of 2006 connected to some 100 artisan groups and had annual sales in excess of $20 million (DeCarlo 2007). Another early alternative trade organization, called SERVE International (Sales Exchange for Refugee Rehabilitation and Vocation), began as an income generation project in 1949 through the sale of wooden clocks carved by refugees in Germany. In 1959, Oxfam, UK, launched the Helping-by-Selling project to import and sell handicrafts (Oxfam 2007). Later that same decade, the first Worldshop opened in the Netherlands. By 2005, there were more than 2,800 Worldshops throughout Western Europe selling mostly fair trade products with annual sales of about $151.8 million (Krier 2005).

Coffee led the rise of fairly traded agricultural products, which followed a path similar to crafts until alternative trade organizations united with other NGOs and traders to create an international certification system in 1988. The emergence of a product certification system allowed the participation of more conventional companies, expanded fair trade markets, and shifted the ratio of global fair trade goods from crafts to foods and beverages. Dutch fair trade advocates were connected to fair trade coffee cooperatives from Guatemala and Mexico by the early 1970s. In the 1980s, two U.S.-based companies, the Equal Exchange workers cooperative and the Thanksgiving Coffee Company, imported Nicaraguan coffee through Canada in protest against the Reagan administration's embargo against the Sandinista government. During the same decade, a large cooperative of small-scale indigenous farmers in southern Mexico called the Unión de Comunidades Indígenas de la Región del Istmo began working with Frans VanderHoff Boersma, a liberation theology inspired Jesuit priest from Holland (VanderHoff Boersma and Roozen 2003). VanderHoff Boersma and others soon identified that a core bottleneck in the fair trade movement was the lack of demand in Northern countries. The small volumes traded through these alternative networks limited revenue generation and community development impacts for small-scale producer organizations.

In response to the problem of how to expand fair trade product demand and distribution without compromising consumer trust, producer groups and Dutch organizers created a *product* certification and labeling system called Max Havelaar. Max Havelaar later united with European and North American NGOs to create Fairtrade Labeling Organizations International (FLO) in 1997. FLO is an international nonprofit multistakeholder association that seeks to establish fair trade standards; support, inspect, and certify disadvantaged producers; and harmonize the fair trade message across the movement (FLO 2007). The combination of activist push, the availability of a credible third-party product certification program (that adheres to ISO 65 standards), consumer demand, and growing public awareness

about the low commodity prices contributed to several large mainstream companies' decision to sell fair trade certified products (Bacon et al. 2008). In an effort to support the alternative aspects within the fair trade movement and distinguish themselves from the large multinational companies, which include Starbucks, Procter and Gamble, and Nestlé, the International Fair Trade Association created a certification for alternative trade *organizations* that implemented fair trade principles throughout their operations.

Impacts, Debate, and Paradox Total sales figures, certification programs, and organizational histories tell us little about the ability of fair trade to deliver on its stated empowerment and sustainable development goals. A primary consideration in assessing fair trade impacts is the fact that as of 2006 only 20 percent of the agricultural goods produced by fair trade certified organizations are sold according to generally accepted fair-trade terms (FLO 2007). Producers generally sell the remaining 80% of their products into lower-paying domestic and international markets. An important percentage is consumed within the household or traded locally. Furthermore, many producers have participated in this market for less than seven years. Most scholars agree that small-scale producers linked to fair trade are better off than producers that lack these connections (Jaffee 2007; Raynolds 2002). Many producers have advanced their sense of collective empowerment by building stronger organizations, and they have also conserved biological and cultural diversity through their farming practices (Bacon et al. 2008). However, the combination of fair trade sales and additional support from allied international development NGOs is not a panacea for eliminating poverty or stopping out-migration even within fair trade organizations (Bacon et al. 2008; Jaffee 2007; Hernández Navarro 2004). The minimum coffee prices are especially important when conventional market prices fall. Cooperatives used their links to fair trade networks to strengthen their capacity and grow as they outcompeted private exporters and transnational trading companies and helped to buffer small-scale farmers from consequences of the coffee

crisis in 1992 and again in 1999 2004 (Bacon et al. 2008). Artisans have generally been able to sustain their crafts and cultures and partially support their livelihoods even as they have improved their organizations' business capacity through direct connections to better markets for their products (Leclair 2002).

Although fair trade has proven itself a useful tool for Southern producer organizations to strengthen their collective influence and for Northern social justice activists leading campaigns to declare fair trade towns and college campuses, it has yet to make significant contributions to international trade and development policy. Fair trade advocates have won voluntary support from the United Nations and the European Commission, and certified fair trade coffee is even served in the U.S. Capitol. No standards have been clearly adopted into global trade regulations, however. Small-scale producers have also used their participation in fair trade to strengthen their alliances and increase their visibility and negotiating power (Renard 2005). For example, the Latin American and Caribbean Network of Small-Scale Fair Trade Producers (CLAC), which represents more than 200,000 producer families, has used its participation in fair trade to win a seat on the FLO board of directors, to gain partial ownership of the certified fair trade system, and to advocate for minimum prices that keep up with inflation and cover the costs of sustainable production (CLAC 2006).

As a voluntary third-party certification system, fair trade contributes to debates about corporate social responsibility and the role of consumers, big business, and more cooperative economies in international sustainable development. One of the certified fair trade system's central paradoxes is that it sets out to achieve social justice and environmental sustainability within the same market that many believe impoverished small producers in the first place. There is also competition with a rapidly expanding array of sustainable product certification programs, such as the Rainforest Alliance and Utz Certified. Although this hybrid approach causes fair trade activists, producer organizations, NGO administrators, and business leaders to live in ideologically

uncomfortable spaces, it may hold the potential to transform market-centered relationships from the inside out.

See also North-South trade; primary products trade; trade and economic development, international

FURTHER READING

Bacon, C. M., V. E. Méndez, S. Gliessman, D. Goodman, and J. A. Fox, eds. 2008. *Confronting the Coffee Crisis: Fair Trade, Sustainable Livelihoods, and Ecosystems in Mexico and Central America.* Cambridge, MA: MIT Press. Uses an action oriented interdisciplinary per spective to analyze the global coffee industry and assess its effects on sustainable community development.

Barratt Brown, Michael. 1993. *Fair Trade: Reform and Realities in the International Trading System.* London: Zed. One of the earlier books by a pioneer scholar and organizer within the fair trade movement.

CLAC. 2006. *Estudio de Costos y Propuesta de Precios para Sostener el Café, las Familias de Productores y Organiza ciones Certificadas por Comercio Justo en América Latina y el Caribe.* Dominican Republic: Assemblea de Co ordinadora Latinoamericana y del Caribe de Pequeños Productores de Comercio Justo. Available at http://www.claccomerciojusto.org/. The most recent study that analyzes the costs of sustainable coffee production and poses a substantial increase in the minimum price for fair trade certified coffee.

DeCarlo, Jacqueline. 2007. *Fair Trade: A Beginner's Guide.* Oxford, UK: One World. An excellent and accessible introduction to fair trade, with a strong focus on crafts.

Fairtrade Labeling Organizations International (FLO). 2007. *Shaping Global Partnerships: FLO Annual Report 2006/2007.* Bonn, Germany: FLO International. An annual report that offers an in depth summary of the size, scope, and accomplishments within the certified fair trade system.

Hernández Navarro, L. 2004. "To Die a Little: Migration and Coffee in Mexico and Central America." Special Report. Silver City, NM: International Relations Center, Americas Program (December). Accessed December 16, 2006, http://www.americaspolicy.org/. Documents the high rates of out migration from coffee growing regions.

Jaffee, Daniel. 2007. *Brewing Justice: Fair Trade Coffee, Sustainability and Survival.* Berkeley: University of

California Press. Combines a well written review of the fair trade coffee industry with a case study among in digenous farmers in Mexico.

Krier, J M. 2005. *Fair Trade in Europe 2005: Facts and Figures on Fair Trade in 25 European countries*. Brussels, Belgium: Fairtrade Labeling Organizations Interna tional, International Fair Trade Association, Network of European World Shops, and European Fair Trade As sociation Fair Trade Advocacy Office. Summarizes the fair trade markets and products (including both crafts and agricultural goods) in Europe.

Leclair, M. S. 2002. "Fighting the Tide: Alternative Trade Organizations in the Era of Global Free Trade." *World Development* 30 (6): 949 58. Summarizes the free trade debates and the role of alternative trade organizations' proposed response.

Oxfam UK, 2007. http://www.oxfam.org.uk/oxfam in action/index.html. Documents Oxfam's pioneering work in humanitarian assistance and fair trade.

Raynolds, Laura T. 2002. *Poverty Alleviation through Par ticipation in Fair Trade Coffee Networks: Existing Research and Critical Issues.* New York: Ford Foundation. An early series of case studies documenting the qualitative impacts of fair trade coffee sales on poverty reduction.

Renard, Marie Christine. 2005. "Quality Certification, Regulation, and Power in Fair Trade." *Journal of Rural Studies* 21: 419 31. An excellent summary of issues of quality, power asymmetries, and debates within the fair trade coffee movement.

VanderHoff Boersma, Franz, and Nico Roozen. 2003. *La Aventura del Comercio Justo: Una Alternativa de Global ización.* Mexico City: El Atajo. A chronicle of history, relationships, and debates in the fair trade movement, written by pioneers of fair trade certification.

CHRISTOPHER M. BACON

■ fear of floating

Since the end of the Bretton Woods system of fixed exchange rates in the early 1970s, the number of countries claiming to be running a flexible exchange rate regime has steadily grown. However, many of these countries appear to actively limit fluctuations in the external value of their national monies. Econo mists Calvo and Reinhart (2002) dubbed this be havior "fear of floating" and argued that this phe nomenon appears pervasive in so-called "emerging markets" (middle-income countries with some ac cess to global financial markets). Indeed, many self-proclaimed "floaters" often accumulate vast war chests of international reserves, which would not be necessary if their currencies were truly floating. In other words, many countries float with "a large life-jacket." Such behavior is puzzling not only because it does not match official pronouncements by pol icymakers, but also because emerging market econ omies are typically buffeted by larger and more frequent external shocks, which in theory necessitate more (not less) exchange rate flexibility.

Why Is Floating So Fearsome? Several inter-locking factors underlie "fear of floating." First, while many industrial countries have operated fairly flexi ble exchange rates quite effectively, they have well-developed and diversified financial systems that are able to minimize real sector disruptions due to transitory exchange rate variations. Most important, industrial countries are able to borrow overseas in their domestic currencies. In contrast, many emerg ing economies are unable to do so, leading to an accumulation of foreign currency debt liabilities that are primarily dollar-denominated and unhedged (i.e., "liability dollarization"). This is commonly re ferred to as the "original sin" problem. In these countries, sharp depreciations in their currencies al ter the domestic currency value of their external debt and therefore the net worth of the economies, with calamitous real sector effects. This in turn may ex plain why exchange rate stability is so important to emerging economies, and why many of these econ omies have an acute "fear of floating."

The inability to borrow overseas in one's own currency is related to the lack of hedging in a number of emerging economies. Even if a country has the ability to hedge, the transaction costs can be too high to make it an attractive option, especially in the short term. In view of this, it may be reasonable to expect some smaller domestic firms in emerging economies to be affected by exchange rate volatility (Bénassy-Quéré, Fontagné, and Lahrèche-Révil 2001). How-

ever, many economists argue that multinational firms can protect themselves from exchange rate fluctuations by maintaining diversified production facilities in different countries and by sourcing from a number of different countries.

Second, policymakers in emerging markets suffer from a chronic lack of credibility. They often have a poor track record in monetary and fiscal policy as a result of, for example, resorting to the inflation tax too often in the past. Consequently, an emerging economy might experience large and frequent shocks to exchange rate expectations or to interest rate risk premiums. A true floater would allow the spot exchange rate to absorb these shocks, while a true pegger would allow the interest rate to adjust in a way to keep the spot exchange rate stable. Governments struck by "fear of floating" allow for some flexibility in both variables, but by and large it is the interest rate that absorbs most of the shock. This might be why various empirical studies have found that emerging economies with officially floating exchange rates have domestic interest rates whose volatility is considerably higher than that in the developed world. Policymakers in these countries might be limiting exchange rate volatility as a way to gain credibility and in order to signal to financial markets their commitment to monetary discipline.

A third reason for "fear of floating" is that small and open economies are relatively more susceptible to exchange rate pass-through effects in domestic prices.

Fourth, countries with flexible regimes have experienced "excessive" volatility over the last few decades. It is difficult to define "excessive" with any precision, but evidence of excessive exchange rate variability comes in a number of forms. For instance, a number of surveys of foreign exchange ("forex") market participants clearly indicate that short-term/high-frequency exchange rate movements are caused by "speculative" or "trend-following" behavior rather than by underlying macroeconomic fundamentals. Indeed, destabilizing speculation is a particular problem in developing countries with thin markets (Indonesia's postcrisis experience since 1999 is a case in point). Of course, even if flexible exchange rates exhibited greater volatility than would be warranted by underlying fundamentals, why might such excessive volatility be of concern? Studies suggest that institutionally fixed exchange rate regimes (i.e., common currency, currency boards, or dollarization) stimulate trade, which in turn boosts income. Proponents of the European Monetary Union (EMU) have used such an argument extensively in support of a single regional currency.

Fifth, some economists argue that access to global financial markets for developing countries is conditioned on currency stability. A sharp depreciation in the nominal exchange rate will often trigger an abrupt pause or even reversal of capital flows into the country (the so-called "sudden stop"). Empirically, such a reversal is associated with a sharp adjustment in the current account (from deficit to surplus), an output contraction, and a collapse in credit ratings.

Finally, there are political reasons behind "fear of floating." Sharp fluctuations in the nominal exchange rate combined with sticky prices translate into unstable relative prices for traded versus nontraded goods. This might cause political disruption in a country in which both the traded and nontraded sectors are large and have powerful lobbies.

Significance Given these explanations for fear of floating, it is reasonable to ask why emerging economies do not fix their currencies explicitly. One reason is that pegging the exchange rate constrains monetary independence. To be sure, some research casts doubt on the extent to which floating regimes in emerging economies provide insulation from foreign interest rate shocks (see Frankel, Schmukler, and Serven 2004; and Hausman, Panizza, and Stein 2001). However, a more recent study using de facto exchange rate regimes for 100 developing and industrial countries between 1973 and 2000 finds that the interest rates of the countries that operated pegged regimes followed the base country far more closely than nonpegs (Shambaugh 2004). There is also some evidence that emerging economies "learn to float" in the sense that as they adopt more flexible exchange rate regimes, they tend to adopt stronger monetary and financial frameworks (Hakura 2005). China appears to be a good example of a country

engaging in this "learning to float" behavior. For example, as the Chinese authorities gradually allow for greater volatility of the currency, they are also putting in place the necessary infrastructure and institutions to ensure that the foreign exchange market functions well (i.e., is liquid) and that agents are able to hedge themselves against volatility. The loss of monetary-policy autonomy can have significant costs that must be carefully considered by policymakers.

See also balance sheet approach/effects; Bretton Woods system; common currency; contagion; currency board arrangement (CBA); discipline; European Monetary Union; exchange rate pass-through; exchange rate regimes; expenditure changing and expenditure switching; foreign exchange intervention; hedging; hot money and sudden stops; impossible trinity; international reserves; original sin; real exchange rate; speculation

FURTHER READING

Bénassy Quéré, A., L. Fontagné, and A. Lahrèche Révil. 2001. "Exchange Rate Strategies in the Competition for Attracting FDI." *Journal of the Japanese and International Economies* 15 (2): 178 98. Examines the impact of exchange rate volatility on FDI.

Calvo, G., and C. Reinhart. 2002. "Fear of Floating." *Quarterly Journal of Economics* 117 (2): 379 408. This paper, along with the following one, launched the debate on "fear of floating."

———. 2005. "Fixing for Your Life." In *Emerging Capital Markets in Turmoil: Bad Luck or Bad Policy?* edited by Guillermo Calvo. Cambridge, MA: MIT Press, 357 402.

Edwards, S. 2006. "The Relationship between Exchange Rates and Inflation Targeting Revisited." NBER Working Paper No. 12163. Cambridge, MA: National Bureau of Economic Research. Examines the impact of exchange rate pass through in developing countries.

Frankel, J., S. Schmukler, and L. Serven. 2004. "Global Transmission of Interest Rates: Monetary Independence and Currency Regime." *Journal of International Money and Finance* 23 (5): 701 33. Examines the empirical validity of the impossible trinity.

Ghosh, A., and R. S. Rajan. 2007. "A Survey of Exchange Rate Pass through in Asia: What Does the Literature Tell Us?" *Asia Pacific Economic Literature* 21 (2): 13 28.

Examines the analytical and empirical literature on ex change rate pass through into inflation in Asian econo mies.

Hakura, D. S. 2005. "Are Emerging Market Countries Learning to Float?" IMF Working Paper No.WP/05/98. Washington, DC: International Monetary Fund. Discusses how the underlying institutional exchange rate and monetary framework of a country changes as economies "learn to float."

Hausmann, R., U. Panizza, and E. Stein. 2001. "Why Do Countries Float the Way They Float?" *Journal of Development Economics* 66 (2): 387 414. An empirical study of "fear of floating," with a special focus on the role of exchange rate pass through and currency mismatches.

Levy Yeyati, E., and F. Sturzenegger. 2005. "Classifying Exchange Rate Regimes: Deeds vs. Words." *European Economic Review* 49 (6): 1603 35. Uses "cluster analysis" to disentangle what countries say versus what they do with regard to exchange rate management.

McKenzie, M. 1999. "The Impact of Exchange Rate Volatility on International Trade Flows." *Journal of Economic Surveys* 13 (1): 71 103. Surveys the literature on the impact of exchange rate volatility on trade.

Rajan, R. S. 2006. "Exchange Rate Policies in Asia since the 1997 98 Crisis." Mimeo (July). George Mason University, Arlington, VA. Examines the empirical trends on exchange rate regimes in Asia since the 1997 98 crisis and the analytical rationale behind the choice of various currency regimes.

Shambaugh, J. 2004. "The Effects of Fixed Exchange Rates on Monetary Policy." *Quarterly Journal of Economics* 119 (1): 300 351. Revisits the empirical validity of the impossible trinity.

SLAVI SLAVOV AND RAMKISHEN S. RAJAN

■ Federal Reserve Board

The Board of Governors of the Federal Reserve System (or the Federal Reserve Board, or "Fed," as it is more commonly known) is the central bank of the United States. It was created by an act of Congress, commonly known as the Federal Reserve Act, in 1913. It is a federal system, with its principal central entity, the Board of Governors, in Washington, DC,

and 12 regional Federal Reserve Banks, one for each of the 12 Federal Reserve Districts, in Boston, New York, Philadelphia, Cleveland, Richmond, Atlanta, Chicago, St. Louis, Minneapolis, Kansas City, Dallas, and San Francisco.

The Federal Reserve System has five main sets of responsibilities: (1) the conduct of monetary policy, (2) bank supervision, (3) addressing systemic risk, (4) acting as the agent for the U.S. government and for foreign official institutions, and (5) operating the payments system of the United States. Generally, the goals of monetary policy, as established by Congress in a number of statutes passed since 1913, include economic growth, price-level stability, high employment, and moderate long-term interest rates.

The Board of Governors is intended to be an independent government agency as well. The board is not funded by congressional appropriations, and governors, once appointed, can be removed from office only "for cause." The Federal Reserve's decisions are not subject to approval by the executive branch, and its funding comes largely from the interest earnings of its portfolio of Treasury debt instruments.

Independence, however, does not mean isolation: members of the Board of Governors testify frequently before congressional committees, and the chairman is required by law to report in February and July of each year. The board has regular contact with the president's Council of Economic Advisors, and the chairman meets regularly with the secretary of the treasury and at times with the president.

Board Responsibility and Composition The Board of Governors has a staff of professionals (about 1,800) who analyze domestic and international economic and financial conditions. The board oversees the Federal Reserve Banks, regulates the operations of the nation's payments system, and administers laws regarding consumer credit protection.

The Board of Governors has seven members, appointed by the president subject to Senate confirmation, for nonrenewable terms of 14 years staggered so that one term would normally expire January 31 of each even numbered year. (A governor appointed to fill the balance of an unexpired term may still be appointed to a full 14-year term.) The chairman and the vice chairman are nominated by the president from the membership of the existing board, subject to Senate confirmation, for four-year terms, renewable through the end of their term as governor. The chairman is the chief executive of the board and is the public face and the spokesperson of the system.

The board and the system are meant to be apolitical: the law governing the board's composition allows for no more than one governor from any Federal Reserve District and calls on the president to make appointments with "due regard to a fair representation of the financial, agricultural, industrial, and commercial interests, and geographical divisions of the country." The staggered terms and fiscal independence provide a substantial measure of insulation from short-run political influence, as a president in the normal course of events would appoint only two governors during each four-year term in office.

Federal Reserve Banks The regional Federal Reserve Banks, which are the operating arms of the Federal Reserve System, include both public and private elements. Each Federal Reserve Bank has a separate board of nine directors, by law chosen from outside the bank. The boards are intended to represent a cross-section of interests within the district: commercial, agricultural, industrial, and public. Some are elected to their posts by the commercial banks in the district that are member banks in the Federal Reserve System, while others are appointed by the Board of Governors.

Conduct of Monetary Policy Responsibility for the conduct of monetary policy lies with the Federal Reserve System, using four mechanisms:

1. Open market operations—the purchase and sale of securities to raise or lower balances held by depository institutions at Federal Reserve Banks;
2. Reserve requirements: the percentage of deposits that depository institutions must hold in reserve as cash or as deposits at Federal Reserve Banks;

3. Contractual clearing balances: amounts that depository institutions agree to hold at Federal Reserve Banks in addition to required reserves; and

4. Discount Window lending: loans by the Federal Reserve Banks to depository institutions.

The Board of Governors is responsible for setting reserve requirements, and it must approve changes in the discount rate initiated by a regional Federal Reserve Bank. Open market operations are the province of the twelve-member Federal Open Market Committee (FOMC), which includes the seven members of the Board of Governors, the president of the New York Federal Reserve Bank, and 4 of the remaining 11 Federal Reserve Bank presidents. Those banks are arranged into four groups of districts   (1) Boston, Philadelphia, and Richmond; (2) Cleveland, Chicago, and Atlanta; (3) St. Louis, Dallas, and Minneapolis; and (4) Kansas City and San Francisco and one Federal Reserve Bank president from each group serves a one-year term in rotation with the others from that group. The nonvoting presidents participate in the Open Market Committee deliberations and evaluation of the economy. Open market operations are conducted by the New York Federal Reserve Bank.

In deploying the tools of monetary policy, the Federal Reserve affects the demand for and supply of deposits held by depository institutions at the Federal Reserve Banks, and thus the federal funds rate, which is the interest rate depository institutions charge one another for overnight loans of deposits at the Federal Reserve banks. The most frequently used   daily, in general   instrument of monetary policy is open market operations, an activity that entails, in simple terms, the purchase (or sale) of U.S. Treasury securities by the Federal Reserve to increase (or decrease) aggregate deposits held by depository institutions at the Federal Reserve, and cause the federal funds rate to fall (or rise). Since 1995, the Federal Reserve has announced its target for the federal funds rate.

The discount rate applies to loans made via the Discount Window, a facility available to virtually all depository institutions that maintain reserves. The rate is usually set about 1 percentage point above the target federal funds rate for what is called primary credit, available to sound institutions for short periods, often overnight. All Discount Window loans are backed by collateral. The facility is an important supplement to open market operations when the supply of balances is less than demand. Further, in cases of natural disaster or other market disruption, the Discount Window provides an avenue for assuring credit availability in the financial system.

Bank Supervision The Federal Reserve Board supervises state-chartered banks that are member banks of the Federal Reserve System, bank holding companies, and the foreign activities of member banks. The board and the Federal Reserve Banks together supervise about 900 state banks and about 5,000 bank holding companies. Bank supervision is also provided by the Comptroller of the Currency (for national banks), and the Federal Deposit Insurance Corporation (for state banks that are not member banks of the Federal Reserve System).

International Responsibilities The chairman has international responsibilities, which include serving as the alternate U.S. member of the board of governors of the International Monetary Fund, a member of the board of the Bank for International Settlements, and a member of the National Advisory Council on International Monetary and Financial Policies. The chairman regularly attends the meetings of central bankers and finance ministers of the G7, and meetings of the Organisation for Economic Co-operation and Development in Paris. The board is also represented at international meetings such as the Finance Ministers' Process of the Asia Pacific Economic Cooperation Forum, meetings of the G20, and meetings of the Governors of Central Banks of the American Continent.

The Federal Reserve's foreign currency transactions are done under the direction of the FOMC and in consultation with the U.S. Treasury, which is responsible for U.S. international financial policy. The manager of the System Open Market Account at the Federal Reserve Bank of New York is the agent for both the Treasury and the Federal Reserve in conducting foreign exchange transactions. Since 1995,

the Federal Reserve has only rarely intervened in foreign currency markets, though in principle it is ready to do so to counter disorderly market conditions. In 1985, when the governments of the five major industrial countries crafted an agreement on exchange rates known as the Plaza Accord, the United States began using foreign exchange market intervention more frequently than it had previously. That intervention, sometimes in concert with other countries' central banks, continued for about a decade, whenever the movement of dollar exchange rates against foreign currencies was deemed excessive.

Publications The board publishes information about the economy and its activities in publications such as the *Federal Reserve Bulletin* (quarterly) and the *Statistical Supplement* (monthly), the semiannual *Monetary Policy Report to the Congress,* and an *Annual Report.* The Federal Reserve Banks provide the board a variety of information from across their districts, used by the board to compile its assessments of the economy and by the FOMC in arriving at its decisions. This information is summarized and made publicly available in a document, popularly known as the Beige Book, issued about two weeks before each FOMC meeting.

See also Asia Pacific Economic Cooperation (APEC); Bank of Japan; deposit insurance; European Central Bank; foreign exchange intervention; inflation targeting; monetary policy rules; money supply; Plaza Accord

FURTHER READING

Friedman, Milton, and Anna Jacobson Schwartz. 1963. *A Monetary History of the United States, 1867 1960.* Princeton, NJ: Princeton University Press. Nontechni cal in its exposition and accessible to the intelligent lay person, this monumental and influential book makes the case that "money matters" using case studies of a large variety of episodes rich in institutional detail, including a major focus on the Great Depression.

Hafer, R. W. 2005. *The Federal Reserve System: An En cyclopedia.* Westport, CT: Greenwood Press. Hafer's volume is aimed at nonspecialists and students, with a wide range of topics including biographical material, and an introduction dealing with the history of central banking in the United States.

Meltzer, Allan H. 2003. *A History of the Federal Reserve.* Chicago: University of Chicago Press. This highly de tailed monetarist perspective on the U.S. central bank, from its inception to the Federal Reserve Treasury Ac cord of 1951, draws from official institutional records and is a complement to the Friedman Schwartz volume.

JAMES LEHMAN

■ Feldstein-Horioka puzzle

There is a widely held perception that the world economy is highly financially integrated in other words, a global finance village exists. However, the accumulated evidence suggests that national markets for physical capital (for example, plant and equipment) are much less integrated than national markets for financial capital (for example, bonds and equities). Martin Feldstein and Charles Horioka, in a well-known article of 1980, were among the first to challenge the view of an international capital market. Using data from 16 industrial countries for the period 1960 74, these authors showed that national investment in physical capital is primarily financed by national saving. Feldstein and Horioka used the following cross-section regression to test market segmentation:

$$I_i = \alpha + \beta S_i + u_i, \qquad (1)$$

where I denotes the ratio of national investment to gross domestic product (GDP), S the ratio of national saving to GDP, and the subscript i the country. If national markets are perfectly segmented, the estimated β is unity. National investment is fully financed by national saving, and capital is not mobile among countries. If, instead, national markets are perfectly integrated, the estimated β is zero. High national investment relative to national saving is financed by capital inflows; low national investment relative to national saving prompts an outflow of capital. Feldstein and Horioka tested and failed to reject the null hypothesis of $\beta = 1$ of zero physical capital mobility.

Interpreting the Feldstein-Horioka Results The Feldstein-Horioka finding instigated a vast empirical literature, which found values of β lower

than unity but larger than zero, especially for the 1980s, but did not disprove the basic conclusion that the markets for physical capital are segmented. Just as important, this literature raised many questions and much criticism on what equation (1) means for capital mobility; three of the most significant criticisms are reviewed here.

The first pertains to the interpretation of parameter β. Consider a classical model in which investment and saving depend exclusively on the real rate of interest, r: r affects negatively I and positively S. In an open economy, an excess of saving over investment implies a net capital outflow or a current account balance surplus, denoted by B. In this model, β is the ratio of the interest sensitivity of investment to the sum of the interest sensitivity of investment, saving, and the current account balance:

$$\beta = I_r/(I_r + S_r + B_r), \qquad (2)$$

where I_r, S_r, and B_r are the slope coefficients of I, S, and B with respect to r (Coakley et al. 1998, 172 73). A simple inspection of equation (2) reveals that complete segmentation of national markets, meaning $\beta = 1$, occurs when both S and B are insensitive to the real interest rate. On the other hand, perfect capital mobility, meaning $\beta = 0$, occurs when either S or B is infinitely sensitive to r. What drives perfect capital mobility: an infinitely elastic saving rate or an infinitely elastic capital account? If we cannot discriminate between alternative causes of high or low β, the test suffers from what is known in jargon as an identification problem. The latter becomes more complex as the underlying models acquire more complexity.

The second criticism concerns the relationship between β and the size of the country. Recall that an excess of saving over investment spills over in a net capital outflow or in a current account surplus in a world of perfect capital mobility. Assume that an unforeseen increase in national saving takes place in a small open economy. The critical assumptions of a small open economy are that its real rate of interest is equal to the world rate of interest and that domestic events cannot influence the world rate of interest. It follows that the world rate of interest and the national investment demand will remain unchanged. The

increase in national saving spills over into larger net capital outflows; the relationship between I and S in equation (1) is such that β is zero. Suppose, instead, that the increase in saving occurs in an economy large enough to influence the world rate of interest. In this case, national I will rise because of the decline in r, and the relationship between I and S in equation (1) is such that β is positive. In sum, β is positively correlated with the size of the economies (Harberger 1980).

The final criticism deals with the use of cross-section data. In the Feldstein-Horioka test and in several subsequent studies, observations are averages of long annual time series. Given that the current account tends to be balanced over long periods of time, I and S tend to equal each other. It follows that the use of time series averages biases the estimate of β toward unity.

Persistence of the "Puzzle" Notwithstanding these criticisms, study after study has confirmed the positive association between investment and saving, to the point that the finding has been elevated to the rank of a "major puzzle" in international macroeconomics (Obstfeld and Rogoff 2001). It is a puzzle because our strong prior, or belief, is that capital is mobile, and our prior has been fed by the evidence on the integration of national markets for financial capital, especially short-term securities issued by industrial countries. Arbitrage tends to quickly eliminate yield differentials. But this parity does not hold for long-term portfolio capital or physical capital (Dooley et al. 1987, 522 23). For national markets of physical capital to be fully integrated, ex-ante real rates of interest in different national markets would have to be equalized; in other words, the following real interest rate parity would have to be satisfied:

$$\begin{aligned} r - r^* = (i - i^* - fp) + (fp - \Delta e) \\ + (\Delta e - \pi + \pi^*) = 0. \end{aligned} \qquad (3)$$

The new symbols in (3) are Δe = the expected depreciation of the home currency and π = the expected rate of inflation. The evidence overwhelmingly rejects (3), and not surprisingly. For (3) to hold, three conditions need to be simultaneously satisfied: covered interest rate parity (the first term in paren-

theses in the equation, the difference in national interest rates adjusted for the premium in the forward exchange market); the forward premium as an unbiased estimate of the expected depreciation (second term, the difference between the forward premium and the expected change in the future spot exchange rate); and expected purchasing power parity (third term, the difference between expected changes in national price levels adjusted for the expected change in the future spot exchange rate). The first of these three conditions, as we have noted, has empirical corroboration for selected currencies and a very narrow set of assets. The second and third conditions fail miserably. In sum, the failure of equation (3) supports the basic contention of Feldstein and Horioka that β is not zero.

Other Home Biases The Feldstein and Horioka finding seems to be consistent with two home biases, one in equities and the other in consumption. The domestic bias in equities is measured relative to the asset diversification predicted by the international capital asset pricing model. Given historical mean returns and variances, the model predicts that the weight of foreign equities should be much higher than the observed weight. The discrepancy between predicted and actual weight remains large even under the assumption of infinite relative risk aversion (Lewis 1999, table 2). The bias could stem from the failure of the capital asset pricing model to predict diversification, or from the failure of purchasing power parity, which is a standard assumption of the international capital asset pricing model, or from the failure of both; there is no way to distinguish between the two. Various attempts to justify the equity home bias have also failed.

The domestic consumption bias is measured relative to the prediction made by a model where markets are complete in the Arrow-Debreu sense and countries diversify risks due to idiosyncratic shocks (Obstfeld and Rogoff 1996, chapter 5). In this setting, the growth rate of domestic consumption is equal to that of foreign consumption. The data clearly refute the implication of complete markets (Lewis 1999, table 1).

In sum, financial capital is more mobile than physical capital. There is an obvious parallel between finance capital and traded goods, and between physical capital and nontraded goods. National borders are just as much an obstacle for the cross-border flows of many goods and services as they are for the flows of physical capital.

See also balance of payments; capital controls; capital mobility; forward premium puzzle; home country bias; interest parity conditions; international capital flows to developing countries; peso problem; purchasing power parity

FURTHER READING

Coakley, Jerry, Flarida Kulasi, and Ron Smith. 1998. "The Feldstein Horioka Puzzle and Capital Mobility: A Review." *International Journal of Finance and Economics* 3: 169 88. Reviews the main issues of the Feldstein Horioka puzzle.

Dooley, Michael, Jeffrey A. Frankel, and Donald J. Mathieson. 1987. "International Capital Mobility: What Do Saving Investment Correlations Tell Us?" *IMF Staff Papers* 34: 503 29. Interprets the Feldstein Horioka evidence as being consistent with a low degree of substitutability for claims on physical capital located in different countries; and physical capital need not be a perfect substitute of financial capital.

Feldstein, Martin, and Charles Y. Horioka. 1980. "Domestic Saving and International Flows." *Economic Journal* 90 (June): 314 29. The original article that shows that countries' investments are highly correlated with their national savings rates.

Harberger, Arnold. 1980. "Vignettes on the World Capital Market." *American Economic Review* 70: 331 37. Argues that capital does not move as freely across national borders as the world capital market view would have it.

Lewis, Karen K. 1999. "Trying to Explain Home Bias in Equities and Consumption." *Journal of Economic Literature* 37 (June): 571 608. Reviews the domestic equity bias and the domestic consumption bias.

Obstfeld, Maurice, and Kenneth Rogoff. 1996. *Foundations of International Macroeconomics.* Cambridge, MA: MIT Press. A graduate level text on international macroeconomics.

. 2001. "The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?" In *NBER Macroeconomics Annual 2000*, edited by Ben Bernanke and Kenneth Rogoff. Cambridge, MA: MIT Press, 339 90. Reviews the major puzzles (or better outcomes inconsistent with widely held views) in inter national macroeconomics.

MICHELE FRATIANNI

■ financial crisis

The financial system is a set of institutions and markets that provides financial intermediation by transferring savings into productive investment. In most developing countries the bulk of financial intermediation has been done via the banking system, with the stock market gaining importance in countries with more advanced institutions. Financial intermediation entails maturity transformation funding a longer-term tangible investment with shorter-term savings. As such, financial intermediation is exposed to financial fragility, in which heightened perceived risk may lead to liquidation, putting the financial system at risk.

Financial crisis refers to a rapid financial disintermediation due to financial panic. In practice, this involves a "flight to quality," where savers attempt to liquidate assets in financial institutions due to a sudden increase in their perceived risk, moving their savings to safer assets, such as foreign currency and foreign bonds in open economies, or currency, gold, and government bonds in closed economies. The ultimate manifestation of financial crises includes bank failures, stock market crashes, and currency crises, occasionally leading to deep recessions.

The economist Hyman Minsky (1964) theorized that financial fragility which is related to the business cycle and to leverage is a typical feature of any capitalist economy. These considerations are at the heart of the large literature propagated by the stock market crash of 1929 and the Great Depression (Bernanke 1995).

Financial Crises and Financial Integration
During the last quarter of the 20th century, observers focused attention on the growing role of international triggers for financial crises an outcome of the collapse of the Bretton Woods system (the post World War II framework for international trade and financial stability), the rapid increase in the importance of emerging markets in the global economy, and the growing financial integration of countries with the global financial system. The resumption of capital flows to developing countries in the early 1990s led to waves of "sudden stops" (the abrupt cessation of foreign capital inflows) and reversals of capital flows, starting with the Mexican crisis of 1994 95, continuing with the Russian and the East Asian crises in the second half of the 1990s, and culminating with the Argentinean meltdown in the early 2000s (Calvo 1998; and Edwards 2004).

Most of the financial crises in the 1990s and early 2000s affected developing and emerging markets, leading to a heated debate regarding their causes and the needed remedies. There is solid evidence that financial opening (that is, the dismantling of capital controls) increases the chance of financial crises. There is more tenuous evidence that financial opening contributes positively to long-run growth. Hence there may be a complex trade-off between the adverse intermediate run and the beneficial long-run effects of financial opening. These findings pose a challenge to policymakers: how to supplement financial opening with policies that would improve this intertemporal trade-off.

To place this issue in a broader context, the debate about financial opening is a reincarnation of the earlier immiserizing growth literature in economics. In particular, while financial opening increases a country's overall welfare when the only distortion is restricting intertemporal trade across countries, financial opening may be welfare-reducing in the presence of other distortions (an economic distortion occurs when an inefficiency prevents an economy from reaching its full potential). An example of such a distortion is moral hazard, which frequently acts as an implicit subsidy to borrowing and investment, ultimately leading to overborrowing and crisis

(McKinnon and Pill 1999; and Dooley 2000). Moral hazard arises when investors believe that they will be bailed out of their bad investments by the taxpayer and, therefore, have little incentive to undertake proper monitoring of their investments. This bailing out may be carried out by the treasury, the central bank, or by international agencies. In these circumstances, the taxpayer subsidizes the investment. A frequent rationale for the bailing out is the "too big to fail" doctrine the fear that allowing large borrowers to go under will trigger a systemic crisis.

Key factors contributing to an exposure to financial crises are balance-sheet features in the form of maturity and currency mismatches between the assets and the liabilities of the banking system, leading to financial fragility. A currency mismatch occurs when residents of the country are not adequately hedged against a change in the exchange rate. This is frequently the case in countries with few foreign assets, serving large external debt denominated in foreign currency, so that a large depreciation generates a large increase in the domestic valuation of the foreign liability, inducing a fall in the economy's net worth, usually accompanied by a large fall in output and insolvencies on the part of firms and banks. Maturity mismatch occurs when the average duration of the liabilities differs from that of the assets. Frequently, banks' liabilities have shorter maturity than banks' assets; hence large withdrawals by consumers may lead to a bank run. Developing countries are more susceptible to balance sheet fragilities and are characterized by debt intolerance: the inability of emerging markets to manage levels of external debt that are manageable for developed, high-income countries (Reinhart, Rogoff, and Savastano 2003).

This literature has led to a spirited debate concerning the wisdom of unrestricted capital mobility between high-income countries and emerging markets. Advocates of financial liberalization in the early 1990s argued that external financing would alleviate the scarcity of savings in developing countries, inducing higher investment and thus higher growth rates. The 1990s experience with financial liberalization suggests that the gains from external financing are overrated the bottleneck inhibiting economic growth has less to do with the scarcity of saving and more to do with other factors, such as the scarcity of good governance (Rodrik 1998; Gourinchas and Jeanne 2003).

Notwithstanding this debate, the strongest argument for financial opening is the pragmatic one. Like it or not, greater trade integration erodes the effectiveness of restrictions on capital mobility (see Aizenman 2004). Hence, for successful emerging markets that engage in trade integration, financial opening is not a question of if, but of when and how. Instead, the hope is that proper sequencing of policies (see McKinnon 1991) and improved coordination will reduce the severity of financial crises, thereby improving the odds of a positive long-run welfare effect of financial opening.

Financial Opening and Financial Crises: The Evidence The recent research has two common themes: it validates empirically the assertion "Goodbye financial repression, hello financial crash" (Diaz-Alejandro 1985). Yet it also has found tenuous evidence that financial liberalization tends to increase growth over time. Both observations suggest an intertemporal trade-off. In the short-run, the fragility induced by financial opening leads frequently to crises. Yet, if these crises force the country to deal with its structural deficiencies, financial opening may induce a higher growth rate in the long run (see Ranciere, Tornell, and Westermann 2005).

Kaminsky and Reinhart (1999) found that problems in the banking sector typically precede a currency crisis; that a currency crisis deepens the banking crisis, activating a vicious spiral; and that financial liberalization often precedes banking crises. Glick and Hutchison (1999) investigated a sample of 90 countries during 1975 97, covering 90 banking crises, 202 currency crises, and 37 twin crises. They found that banking and twin crises have occurred mainly in developing countries, and their number increased in the 1990s. Twin crises are mainly concentrated in financially liberalized emerging-market economies. The costs of these crises are substantial currency (banking) crises are very costly, reducing output by about 5 percent 8 percent (8 10 percent)

over a two- to four-year year period (Hutchison and Noy 2005).

A useful survey of financial liberalization is found in Williamson and Mahar (1998), which focused on 34 countries that undertook financial liberalization between 1973 and 1996. Overall, the authors found a mixed record of financial liberalization the gains are there, but the liberalization carries the risk of a financial crisis. Financial liberalization has yielded greater financial depth and increased efficiency in the allocation of investment. Yet it has not brought the boost in saving. The main recommendations emerging from their study are akin to those in Hellman, Murdock, and Stiglitz (2000) start with macroeconomic stabilization, improve bank supervision, while delaying capital account convertibility to the end of the process. Maintaining high spreads may be needed in a transition until banks are able to work off the legacy of bad debt inherited from the period of financial repression, preventing moral hazard associated with a "gamble for resurrection."

The overall effect of financial opening on growth remains debatable. Rodrik (1998) failed to detect any positive effects of financial opening on investment, growth, and inflation. Bekaert, Harvey, and Lundblad (2001) found that equity market liberalizations, on average, lead to a 1 percent increase in annual real economic growth over a five-year period. The investment/gross domestic product ratio increases postliberalization, with the investment partially financed by foreign capital, inducing worsened trade balances. The liberalization effect is enhanced by a large secondary school enrollment, a small government sector, and an Anglo-Saxon legal system.

In summary, recent financial crises affecting developing countries are the outcome of financial fragilities, reflecting the downside of growing financial integration. The challenge is mitigating the pain in ways that enhance growth and economic welfare.

See also asymmetric information; banking crisis; Bretton Woods system; capital flight; currency crisis; deposit insurance; financial liberalization; financial repression; international reserves; lender of last resort; original sin; sequencing of financial sector reform

FURTHER READING

Aizenman, Joshua. 2004. "Financial Opening and Development: Evidence and Policy Controversies." *American Economic Review* 94 (2): 65 70. Greater trade integration erodes the effectiveness of restrictions on capital flow. Thus, for a country that engages in trade integration, financial opening is not a question of if, but of when and how.

Bekaert, G, C. Harvey, and R. Lundblad. 2001. "Does Financial Liberalization Spur Growth?" NBER Working Paper No. 6724. Cambridge, MA: National Bureau of Economic Research. In a large sample of countries since 1980, financial liberalization leads to a 1 percent increase on average in a country's annual growth rate over a five year period.

Bernanke, Ben S. 1995. "The Macroeconomics of the Great Depression: A Comparative Approach." *Journal of Money, Credit, and Banking* 27 (1): 1 28. Monetary shocks as aggregate demand force behind the Depression; the induced financial crisis and sticky nominal wages are supply mechanisms by which nominal shocks had real effects.

Calvo, Guillermo A. 1998. "Capital Flows and Capital Market Crises: The Simple Economics of Sudden Stop." *Journal of Applied Economics* 1 (1): 35 54. Articulates mechanisms through which a sudden stop in international credit flows may bring about financial and balance of payments crises, and examines factors triggering sudden stops.

Catao, Luis, and Sandeep Kapur. 2006. "Volatility and the Debt Intolerance Paradox." *IMF Staff Papers* 53 (2): 195 218. Differences in macroeconomic volatility account for debt intolerance of developing countries the ability to borrow is constrained by higher default risk that volatility engenders.

Chang R., and A. Velasco. 1999. "Financial Crises in Emerging Markets." *Economic Review* (Federal Reserve Bank of Atlanta) 84 (2): 4 17. A bank collapse multiplies the harmful effects of an initial shock. Under fixed exchange rates, a run on banks becomes a run on the currency if the central bank acts as a lender of last resort.

Demigüc Kunt A., and E. Detragiache. 1998. "Financial Liberalization and Financial Fragility." IMF Working Paper No. 98/83. Washington, DC: International Monetary Fund. Banking crises are more likely to occur

in liberalized financial systems; the impact of financial liberalization on banking sector fragility is weaker where the institutional environment is strong.

Diaz Alejandro, C. 1985. "Goodbye Financial Repression, Hello Financial Crash." *Journal of Development Eco nomics* 19 (1/2): 1 24. Unintended consequences of fi nancial liberalization: intrinsic imperfections in financial markets imply that lifting of financial repression leads frequently to financial crash.

Dooley, Michael. 2000. "A Model of Crises in Emerging Markets." *The Economic Journal* 110 (460): 256 72. Variety of shocks generate capital inflows to emerging markets followed by successful and anticipated specu lative attacks, liquidating reserve assets accumulated as self insurance.

Edwards, Sebastian. 2004. "Financial Openness, Sudden Stops, and Current Account Reversals." *American Eco nomic Review* 94 (2): 59 64. Restricting capital mobility does not reduce the probability of experiencing a rever sal; flexible exchange rate regimes are better able to ac commodate shocks stemming from a reversal.

Glick R., and M. Hutchison. 1999. "Banking and Currency Crises: How Common Are Twins?" In *Financial Crises in Emerging Markets*, edited by Reuven Glick, Ramon Moreno, and Mark Spiegel. New York: Cambridge University Press, 35 69. Crisis phenomenon is most common in financially liberalized emerging markets; banking crises are a leading indicator of currency crises in emerging markets. The converse does not hold.

Gourinchas, Pierre Olivier, and Olivier Jeanne. 2003. "The Elusive Gains from International Financial Integration." NBER Working Paper No. 9684. Cambridge, MA: National Bureau of Economic Research. Welfare gains from switching from financial autarky to perfect capital mobility in neoclassical models are negligible relative to the welfare gain of a take off in domestic productivity.

Hellmann, F. Thomas, Kevin C. Murdock, and Joseph E. Stiglitz. 2000. "Liberalization, Moral Hazard in Bank ing, and Prudential Regulation: Are Capital Require ments Enough?" *American Economic Review* 90 (1): 147 65. Competition can undermine prudent bank behavior in the presence of moral hazard. Pareto effi cient outcomes can be achieved by adding deposit rate controls as a regulatory instrument.

Hutchison, Michael, and Ilan Noy. 2005. "How Bad Are Twins? Output Costs of Currency and Banking Crises." *Journal of Money, Credit, and Banking* 37 (4): 725 52. Currency (banking) crises are very costly, reducing output by about 5 percent to 8 percent (8 10 percent) over a two to four year period, no additional feedbacks associated with twin crises.

Kaminsky, Graciela L., and Carmen M. Reinhart. 1999. "The Twin Crises: The Causes of Banking and Balance of Payments Problems." *American Economic Review* 89 (3): 473 500. Problems in the banking sector typically precede a currency crisis; currency crisis deepens the banking crisis, activating a vicious spiral; financial lib eralization often precedes banking crises.

McKinnon, Ronald. 1991. *The Order of Economic Liberali zation: Financial Control in the Transition to a Market Economy.* Baltimore: Johns Hopkins University Press. Financial control and macroeconomic stability are more critical to a successful transition than is any crash pro gram to privatize state owned industrial assets and the banking system.

McKinnon, Ronald, and Huw Pill. 1999. "Exchange Rate Regimes for Emerging Markets: Moral Hazard and In ternational Overborrowing." *Oxford Review of Economic Policy* 15 (3): 19 38. Investigates the overborrowing syndrome. A "good" exchange rate peg stabilizes the domestic economy while limiting moral hazard in the banking system.

Minsky, M. Hyman. 1964. "Longer Waves in Financial Relations: Financial Factors in the More Severe De pressions." *American Economic Review* 54 (2): 324 35. Firms' need to finance investment spending and capital externally explains longer and deeper swings in eco nomic experience.

Ranciere Romain, Aaron Tornell, and Frank Westermann. 2005. "Systemic Crises and Growth." NBER Working Paper No. 11076. Cambridge, MA: National Bureau of Economic Research. Financial liberalizations facili tate risk taking, increasing leverage and investment, leading to higher growth, but also to a greater incidence of crises.

Reinhart, C., K. Rogoff, and M. Savastano. 2003. "Debt Intolerance." *Brookings Papers on Economic Activity* 1: 1 74. Debt intolerance is the duress many emerging mar kets experience at debt levels that are manageable by

advanced country standards, and is linked to default and inflation history.

Rodrik, Dani. 1998. "Who Needs Capital Account Con vertibility?" In *Should the IMF Pursue Capital Account Convertibility? Essays in International Finance No. 207* (May), edited by Peter Kenen. Princeton, NJ: Princeton University Press, 55 65. The magnitude of recent crises is not justified by changes in the fundamentals of the affected economies; capital account liberalization is not necessary as a tool for growth.

Williamson, John, and Molly Mahar. 1998. "A Survey of Financial Liberalization." *Princeton Essays in Interna tional Finance No. 211.* Princeton, NJ: Princeton Uni versity Press. Evidence that financial liberalizations lead to financial deepening and increase the efficiency of investment allocation, but frequently spawn financial crisis.

JOSHUA AIZENMAN

■ financial integration

See capital mobility

■ financial liberalization

Financial liberalization, or *financial reform*, in the most general sense, refers to the transition away from a financial system characterized by state intervention and ownership and toward a more market-oriented system. Not only does financial liberalization need to be considered in discrete historical phases that differ by region and level of development, it also can and should be understood as it applies to the domestic and international spheres. Sometimes this distinc tion boils down to the difference between banking reform and capital market reform. Outward, or ex ternally oriented, financial liberalization involves lifting state controls on the flow of finance between the country in question and international financial markets, including allowing for competition from international banks operating within the country. Internally oriented financial liberalization focuses on removing restrictions on competition and other business practices within the domestic financial markets.

Internally Oriented Controls and Liberal ization One of the most common forms of restric tions placed on domestic banks in both developed and developing countries has been interest rate ceil ings that is, restrictions on the interest rate banks are allowed to pay depositors in order to compete for their deposits, and/or limits on the interest rate banks are allowed to charge borrowers on their loans. De regulation or liberalization in this case entails re moving such ceilings.

Another form of interest rate regulation takes a more microeconomic form. Governments can set different interest rate ceilings depending on the economic sector of the loan. In this way, govern ments gain control over where banks choose to direct financial resources and who has access to credit. This is sometimes referred to as sector-specific allocation of finance, which has been much more common among newly industrializing economies (NIEs) in Asia than in Latin America or advanced industrial countries. Another method of sector-specific alloca tion is the manipulation of marginal reserve re quirements on bank deposits. For example, mone tary officials could control the allocation of private investment by altering the level of reserves private banks are required to hold in the central bank. Sometimes the reserve requirements would be ad justed depending on the sector of the economy to which loans were destined. (For a more detailed discussion of how this worked in Mexico, see Auer bach 2001). Especially in Asia, directed finance was part of a broader industrial policy strategy of state-led industrialization.

The most direct way that governments have controlled the flow of domestic finance, particularly in developing countries, is through ownership of banks. Thus privatization, that is, the process of selling off these banks to the private sector, often becomes the cornerstone of financial liberalization for the domestic market. Even after the domestic banking sector has been privatized, however, gov ernments often and to varying degrees retain some control over where private sector banks direct their

financial resources. This phenomenon, often called "window guidance," was mostly practiced in Asian economies by heavily interventionist developmental states such as South Korea and Japan that were unwilling to give up complete control over the allocation of finance despite the intense pressure to liberalize.

Another of the mechanisms by which governments have retained control over privatized banking sectors is through the authority to charter new banks. While this authority, much like the authority to print money, falls naturally under the purview of the state and would not necessarily be considered interventionist, the level of restrictions on new charters and the form that those restrictions take do fall under the umbrella of state intervention. Thus when restrictions on chartering are lifted altogether, or when preferential charter policies are replaced with neutrally applied charters that treat foreign and domestic owners alike, these policy reforms would be considered a form of financial liberalization.

Restrictions on banking competition have taken multiple and specific forms often related to particular national histories. In the United States, for example, legislation once mandated that banks could not take deposits in more than one state and that banks could not directly own stock in industrial firms. The repeal of this legislation in 1999 represented for the United States one of the last steps in a financial reform process that began in the 1970s with the lifting of interest rate ceilings.

External Financial Liberalization Externally oriented liberalization aims to remove restrictions of financial flows or price controls between the domestic financial market and international finance. The removal of capital controls in advanced industrial countries took place mostly by the 1970s primarily because of the ineffectiveness of these controls in the context of new technologies that allowed banks to transfer capital internationally at the touch of a button. Yet there have been notable exceptions to this rule. Chile imposed a significant tax on capital outflows, known as the "Chile tax," in the 1990s that effectively controlled the outflow of capital.

The second major type of external liberalization came in the form of allowing the entry of foreign banks into domestic banking markets that had been previously restricted to domestic ownership. This type of liberalization has been slow and piecemeal given the politically charged role that banking plays in most economies.

Liberalizing exchange rates has resulted in part from the pressures of increasingly mobile international capital. Most experts have agreed that managing exchange rates has become increasingly difficult and risky given the tendency for exchange rate speculation in international currency markets. While governments have continued to see the exchange rate as a key economic variable affecting trade competitiveness and foreign investment flows, their ability to manipulate that variable waned considerably in the 1990s.

Waves of Financial Liberalization Historically and globally, financial liberalization has come in waves. The 1970s saw a major wave of financial reforms among advanced industrial countries. Some authors frame the wave of market-oriented reforms among advanced industrial countries as part of the breakdown of the Bretton Woods system, which promoted state intervention through tight control over international financial flows. The end of the Bretton Woods system resulted in an immediate shift from pegged, or fixed, to floating exchange rates for most advanced industrial countries. Most attempted to maintain capital controls, but rapidly increasing financial flows aided by new computer technologies rendered these policies increasingly costly and difficult to monitor. Capital controls in the United States were permanently removed in 1974, and in Britain in 1979. The last European countries to lift controls were Portugal and Ireland in the early 1990s.

In contrast to the advanced industrial country experience, developing countries, especially the NIEs, maintained more state interventionist or what some have called "repressed" financial systems until the 1980s. Here the catalyst for the wave of reforms came as a result of the Latin American debt crisis. Again, interpretations of this wave of financial liberalization differ widely among proponents and

opponents of financial liberalization. Proponents are apt to suggest that the debt crisis was the natural result of years of inefficient financial intervention or repression of financial markets. Opponents of financial liberalization, on the other hand, argue that the debt crisis put developing countries between a rock and a hard place: under international pressure from the United States and the International Monetary Fund (IMF), they had little choice but to liberalize their financial markets or face being permanently cut off from international capital flows. As a result, most Latin American countries and some Asian countries, to varying degrees, lifted interest rate ceilings, stopped using marginal reserve requirements to direct finance to specific sectors, loosened capital controls, and began privatizing the banking sector as part of the first rounds of financial reforms in the 1980s.

With the 1990s came a new set of financial crises that precipitated a new round of financial liberalization, this time focused on external capital markets and domestic banking. Lively debate over the pros and cons of financial liberalization has arisen among economists and political scientists as well. Particularly after the Asian financial crisis of 1997, proponents of financial liberalization blamed insufficient liberalization in the previous decade as the primary cause, while opponents blamed overly rapid and extensive liberalization, particularly the liberalization of external capital flows (see, e.g., Furman and Stiglitz 1998; and Rodrik 1996).

Most of the political economy literature on financial liberalization has focused on the causes of liberalization without differentiating between perverse and benign forms of liberalization. A growing political economy literature now rejects the notion that financial liberalization is either inherently beneficial or harmful, but instead describes the political conditions under which liberalization will promote growth and efficiency as opposed to perverse outcomes.

See also banking crisis; Bretton Woods system; capital controls; capital flows to developing countries; convertibility; currency crisis; exchange rate regimes; financial crisis; Washington consensus

FURTHER READING

Auerbach, N. 2001. *States, Banks, and Markets: Mexico's Path to Financial Liberalization in Comparative Perspective.* Boulder, CO: Westview Press. Discusses Mexico's experience with financial liberalization.

Furman, J., and J. Stiglitz. 1998. "Economic Crises: Evidence and Insights from East Asia." *Brookings Papers on Economic Activity No. 2*, 1 114.

Rodrik, D. 1996. "Understanding Economic Policy Reform." *Journal of Economic Literature* 34: 9 41. The Rodrik and Furman and Stiglitz articles argue that other economists oversold the benefits of the free movement of financial flows internationally as part of the benefits of financial liberalization.

Smith, P. H., K. Horisaka, and S. Nishijima. 2003. *East Asia and Latin America: The Unlikely Alliance.* Lanham, MD: Rowman and Littlefield. Articles in this book offer an interesting comparison of directed credit, industrial policies, and financial liberalization in Asia versus Latin America.

Williamson, J., and M. Mahar. 1998. "A Survey of Financial Liberalization." *Essays in International Finance No. 211.* Princeton: Princeton University, Department of Economics, International Finance Section. An excellent overview of the aspects of financial liberalization, including sequencing issues.

NANCY NEIMAN AUERBACH

■ financial repression

Financial repression refers to the notion that a set of government regulations, laws, or other nonmarket restrictions prevents an economy's financial intermediaries such as banks and security markets from functioning at their full capacity. The policies that cause financial repression include interest rate ceilings, liquidity ratio requirements, high bank reserve requirements, capital controls, restrictions on market entry into the financial sector, credit ceilings or restrictions on directions of credit allocation, and government ownership or domination of banks. Economists have commonly argued that financial repression prevents the efficient allocation of capital and thereby impairs economic growth.

The economists McKinnon (1973) and Shaw (1973) were the first to explicate the notion of financial repression. While theoretically an economy with an efficient financial system can achieve growth and development through efficient capital allocation, McKinnon and Shaw argue that historically, many countries, including developed ones but especially developing ones, have restricted competition in the financial sector with government interventions and regulations. According to McKinnon and Shaw's argument, a repressed financial sector discourages both saving and investment because the rates of return are lower than what could be obtained in a competitive market. In such a system, financial intermediaries do not function at their full capacity and fail to channel saving into investment efficiently, thereby impeding the development of the overall economic system.

Rationale for and Types of Financial Repression The control of fiscal resources is a key reason for the implementation of financially repressive policies by governments. By having a direct control over the financial system, the government can funnel funds to itself without going through legislative procedures and more cheaply than it could when it resorts to market financing. More specifically, by restricting the behavior of existing and potential participants in the financial markets, the government can create monopoly or captive rents (that is, economic profits that can arise from market imperfection) for the existing banks and also tax some of these rents so as to finance its overall budget, which may be economically attractive to both the existing banks and the government. Existing banks may try to collude with one another and to interrupt possible liberalization policies as long as they are guaranteed their collective monopoly position in the domestic market.

In some countries, governments require banks to meet high reserve ratios and seize some of these reserves as government revenues. Because reserves earn no interest, reserve requirements function as an implicit tax on banks and also restrict banks from allocating a certain portion of their portfolios to productive investments and loans. When high reserve ratios are required, the lending and borrowing rate spread must widen to incorporate the amount of no-interest reserves, which can reduce the amount of funds available in the financial market. If high reserve requirements are combined with interest ceilings and protective government directives for certain borrowers, savers who are usually unaware of the requirement policy become the main "taxpayers" because they face reduced rates of interest on their savings. Inflation can aggravate the "reserve tax" because it reduces the real rates of interest. Thus high reserve requirements make the best use of the government's monopolistic power to generate seigniorage revenue (the revenue monetary authorities can create by issuing currency) as well as to regulate reserve requirements. A variant of this policy includes required liquidity ratios banks are required to allocate a certain fraction of their deposits to holding government securities that usually yield a return lower than could be obtained in the market.

Governments often impose a ceiling on the interest rate banks can offer to depositors. Interest ceilings function in the same way as price controls, and thereby provide banks with economic rents. Like high required reserve ratios, those rents benefit incumbent banks and provide tax sources for the government, paid for by savers and by borrowers or would-be borrowers. The rents borne by the interest ceiling reduce the number of loans available in the market the real interest rates on loans and deposits are higher and lower, respectively, thereby discouraging both saving and investment. In return for allowing incumbent banks to reap rents, the government often requires banks to make subsidized loans to certain borrowers for the purpose of implementing industrial policy (or simply achieving some political goals). Interest ceilings in high inflation countries can victimize savers because high inflation can make the real interest rates of return negative.

Financial repression also takes the form of government directives for banks to allocate credit at subsidized rates to specific firms and industries to implement industrial policy. Forcing banks to allocate credit to industries that are perceived to be strategically important for industrial policy ensures stable provision of capital rather than leaving it to

decisions of disinterested banks or to efficient securities markets. It is also more cost effective than going through the public sector's budgetary process. Government directives and guidance sometimes include detailed orders and instructions on managerial issues to financial institutions to ensure that their behavior and business are in line with industrial policy or other government policies. The Japanese Ministry of Finance (MOF) is a typical example of government's micromanagement of the financial industry. Till the 1980s, for example, the MOF segmented the banking industry into different types of banks and specified what kind of borrowers and depositors banks in each type can deal with, along with other specific directives and "administrative guidance" about the banking business allowed for each type of banks. The extreme example of direct state control of banks is the nationalization of banks as was observed in Mexico in the 1980s, when the government took control of all the banks to secure public savings.

Capital controls are restrictions on the inflows and outflows of capital, and are also financially repressive policies. Despite their virtues, the use of capital controls can involve costs. Because of their uncompetitive nature, capital controls increase the cost of capital by creating financial autarky (that is, restricting access to cheaper capital available outside the domestic market), limit domestic and foreign investors' ability to diversify their portfolios, and help inefficient financial institutions to survive by protecting them from international competition.

Impacts of Financial Repression Because financial repression leads to inefficient allocation of capital, high costs of financial intermediation, and lower rates of return to savers, it is theoretically clear that financial repression inhibits growth (Roubini and Sala-i-Martin 1992). The empirical findings on the effect of removing financial repression, that is, financial liberalization, on growth support this view. These findings suggest that liberalization spurs growth through various channels.

The possible negative effect of financial repression on economic growth does not automatically mean that countries should adopt a laissez-faire stance on financial development and remove all regulations and controls that create financial repression. Many developing countries that liberalized their financial markets experienced crises partly because of the external shocks that financial liberalization introduces or amplifies. Indeed, financial liberalization can create short-term volatility despite its long-term gains (Kaminsky and Schmukler 2002). Also, because of market imperfections and information asymmetries (in which one party to a financial transaction has better information than the other), removing all public financial regulations may not yield an optimal environment for financial development. An alternative to a financially repressive administration would be a new set of regulations to ensure market competition as well as prudential regulation and supervision.

See also asymmetric information; banking crisis; capital controls; capital flight; convertibility; financial liberalization; money supply; seigniorage

FURTHER READING

Ariyoshi, Akira, Karl Habermeier, Bernard Laurens, Inci Otker Robe, Jorge Iván Canales Kriljenko, and Andrei Kirilenko. 2000. "Capital Controls: Country Experiences with Their Use and Liberalization." IMF Occasional Paper 190 (May). Washington, DC: International Monetary Fund. Presents an overview of capital controls by discussing the role of capital controls and the current trend of capital account liberalization.

Beim, David, and Charles W. Calomiris. 2001. *Emerging Financial Markets*. New York: McGraw Hill. Discusses factors that contribute to the development of financial markets, especially those in emerging market countries. One of the chapters discusses the effect of financial repression.

Claessens, Stijin. 2005. "Finance and Volatility." In *Managing Economic Volatility and Crises: A Practitioner's Guide*, edited by J. Aizenman and B. Pinto. Boston: Cambridge University Press, 213 80. Reviews the academic literature on the determinants of financial markets; the role of financial markets in economic development; the potential impact of domestic financial deregulation; and the effect of financial volatility on the economy.

Kaminsky, Graciela, and Sergio Schmukler. 2002. "Short Run Pain, Long Run Gain: The Effects of Financial Liberalization." World Bank Working Paper No. 2912. Washington, DC: World Bank. Examines empirically the short and long run effects of financial liberalization on capital markets and finds financial liberalization can lead to pronounced boom bust cycles in the short run, but more stable markets in the long run.

McKinnon, Ronald I. 1973. *Money and Capital in Economic Development.* Washington, DC: Brookings Institution. A seminal work along with Shaw (1973) in which the author defines financial repression and identifies it in the development of financial markets in countries.

Montiel, Peter. 2003. *Macroeconomics in Emerging Markets.* Cambridge: Cambridge University Press. Provides mac roeconomic theories and models especially for emerging market countries. A few chapters in the book present academic discussions about financial repression and the impact of financial liberalization.

Roubini, N., and X. Sala i Martin. 1992. "Financial Re pression and Economic Growth." *Journal of Develop ment Economics* 39: 5 30. One of the first papers that empirically analyzes the negative effects of financial re pression on economic growth.

Shaw, Edward. 1973. *Financial Deepening in Economic De velopment.* New York: Oxford University Press. Along with McKinnon (1973), this book is a seminal work that defines financial repression and examines the effect of financial development on economic development.

HIRO ITO

■ financial services

A nation's financial system provides an essential infrastructure for the functioning of the entire economy. A sound and efficient financial system is therefore imperative for economic growth and development. A sound financial system also increases the resiliency of a nation's economy, thereby helping it to withstand external shocks such as movements in exchange rates or a major increase in global interest rates.

International trade in financial services in combination with enhanced prudential regulation and supervision and other basic structural reforms can play an important role in helping countries build financial systems that are more competitive and efficient and therefore more stable. Trade in financial services can enhance capital market efficiency and improve the quality, availability, and pricing of financial services. Especially when a foreign financial firm operates through a commercial presence in a host country, trade in financial services can stimulate innovation through the dissemination of new technologies, know-how, and skills. Trade in financial services can also promote the use of international good practices for example, with regard to risk management and disclosure of financial information.

Trade liberalization and strengthening domestic financial systems the two major dimensions of the international framework for financial services are complementary and mutually reinforcing. Indeed, in order to obtain the maximum benefits of liberalization while minimizing the risks, a host country must ensure adequate prudential regulation and supervision and put in place other elements of the institutional and legal structure necessary for a sound financial system. International work on strengthening domestic financial systems takes place in a variety of forums, ranging from the International Monetary Fund (IMF) to specialized bodies such as the Basel Committee on Banking Supervision.

Liberalization of trade in financial services is dealt with in the World Trade Organization (WTO) under the General Agreement on Trade in Services (GATS) and also in regional and bilateral trade arrangements.

International Trade in Financial Services Major financial firms now provide a wide range of services to customers in other countries. The GATS defines a financial service as "any service of a financial nature . . . including insurance and insurance-related services and all banking and other financial services" and provides a list of activities included in this definition. The list includes commercial banking activities such as lending and deposit-taking; investment banking activities, such as underwriting securities and advising on mergers and acquisitions;

trading activities, that is, brokering and dealing in securities and other financial instruments; and asset-management activities, including management of mutual funds and pension funds. Other noninsurance financial services on the list comprise financial information and data processing services; investment advisory services; payment and money transmission services, including credit cards; settlement and clearing for financial assets; and financial leasing. Insurance services on the GATS list include "direct" life and nonlife insurance, that is, insurance sold to the public and to noninsurance businesses; reinsurance and "retrocession" (reinsurance by reinsurers); and commercial insurance. The list also includes services auxiliary to insurance such as actuarial, risk assessment, and claims settlement services.

"Trade in services" as defined in the GATS covers both foreign direct investment (FDI) as well as cross-border trade, which is broadly analogous to trade in goods. FDI includes services provided through a "commercial presence" in a host country, such as a branch or subsidiary of a foreign bank. The GATS definition of commercial presence also includes representative offices, which are basically a marketing device and do not constitute FDI. Trade in services as defined in the GATS also includes services provided in a host country by "natural persons" (individuals) who are residents of another country. In the financial services sector, however, individuals usually provide services as employees of financial firms; exceptions are financial advisers and, in insurance, independent sales intermediaries.

A cross-border financial services transaction could be carried out in a number of ways. For example, a representative of a foreign bank might visit the customer's country to arrange a loan. Alternatively, the customer might travel abroad to visit the office of the foreign bank. Or, the transaction might take place via telephone, fax, or, increasingly, the Internet, which in this context is another technological means of carrying out a transaction. Although the GATS distinguishes between cross-border services provided to nonresidents "from" and "in" the country of the service supplier (referred to as "cross-border supply" and "consumption abroad," respectively), for financial services the line dividing these two modes of supply is not always clear.

Wholesale financial services   that is, services provided to "sophisticated" customers such as corporations and institutions, other financial services firms, and, under some definitions, wealthy individuals   are provided internationally both through FDI and across borders. Electronic transactions play a major role in the cross-border provision of wholesale financial services, including traditional financial services as well as newer types of services designed to facilitate business-to-business e-commerce activities. When wholesale financial services are provided through FDI, foreign financial firms typically prefer to use the branch form of organization. Branches, unlike subsidiaries, are not separately incorporated in the host country and, unless restricted by host-country regulations, operate using the firm's consolidated worldwide capital.

Retail financial services are still provided internationally primarily through FDI. Although the provision of some financial services over the Internet and through Web-enabled technologies, such as mobile telephony, is expanding dramatically *within* a number of countries, the cross-border provision of financial services to retail customers through these methods is still in its infancy. The lack of widespread development of cross-border retail banking and other financial services   whether through electronic transactions or more traditional methods   reflects host-country regulatory requirements aimed at ensuring adequate consumer protection, consumer preferences, and tax considerations. Some countries actually require the establishment of a commercial presence to provide retail financial services. Perhaps even more important, consumers may prefer dealing with a local commercial presence, particularly because redress against a local establishment is usually readily available through the domestic legal system. In addition, in a number of countries, consumers receive more favorable tax treatment on financial products that are provided through locally incorporated entities.

Strengthening Domestic Financial Systems
The financial services sector has an elaborate and intensively used framework of international forums that are used to address overall financial and regulatory policy issues, to promote cooperation and co-ordination among supervisors, to set voluntary but widely accepted international minimum standards and codes of good practices, and to provide "surveillance" of domestic financial systems. This international framework, which has been constructed over three decades and is still evolving, is a response not only to the internationalization of banking and other financial activities but also to the special characteristics of the financial sector, especially the phenomenon of "systemic risk" whereby problems with one financial firm can be transmitted to unrelated financial firms, both within and beyond a single country.

International minimum standards and codes of good practices for sound financial systems have been established in three broad areas: (1) transparency of macroeconomic policy and data; (2) institutional and market infrastructure, which covers insolvency, corporate governance, accounting, auditing, market integrity and functioning, and payment and settlement systems; and (3) prudential regulation and supervision, which covers both financial firms and regulatory and supervisory systems. Examples are the Code of Good Practices on Transparency in Monetary and Financial Policies promulgated by the IMF, the Principles of Corporate Governance issued by the Organisation for Economic Co-operation and Development (OECD), and the Core Principles for Effective Banking Supervision developed by the Basel Committee on Banking Supervision.

The international work on strengthening domestic financial systems includes financial sector surveillance conducted by the IMF and the World Bank under a joint Financial Sector Assessment Program (FSAP). The FSAP is designed to provide a "comprehensive health check-up of a country's financial sector" by identifying the strengths and vulnerabilities of a country's financial system, determining whether effective risk-management techniques are being used, and evaluating the observance of internationally accepted minimum standards and codes. A continuing challenge for the IMF is to integrate its financial sector surveillance more effectively into its so-called Article IV surveillance of the economic policies of its member countries. Technical assistance for a country that wishes to enhance its regulatory and supervisory capabilities is available under the auspices of the IMF and World Bank, or through programs established by national regulatory authorities.

Liberalization in the GATS The GATS, which was negotiated in the Uruguay Round (1986–94), is the first global trade agreement to cover financial and other services. The GATS has two major components. First, it establishes overall rules and disciplines for trade in services, both in the main text and in various annexes; the Annex on Financial Services deals with rules specific to financial services. Second, the GATS contains each WTO member's schedule of specific commitments and list of exemptions from the most-favored-nation (MFN) obligation. The results of the financial services negotiations—which were not concluded until December 1997, several years after the end of the Uruguay Round—are incorporated into these schedules and lists by the Fifth Protocol to the GATS. Although the expression "agreement on financial services" is widely used to refer to the results of the 1997 negotiations or, more broadly, to the GATS as it applies to financial services, from a legal point of view, a separate multilateral agreement for financial services does not exist.

The inclusion of the financial services sector in a multilateral trade agreement was a major milestone. Because of the special characteristics and sensitivity of the financial sector—in particular, the role of banks in monetary and payment systems and the phenomenon of systemic risk—finance officials in a number of countries were concerned about allowing financial sector issues to fall within the domain of trade officials and the multilateral trading system. Financial regulators made it clear that the inclusion of financial services in the GATS would be unacceptable without a specific carve-out from the obligations of the agreement for prudential

435

measures. The regulators emphasized that such a carve-out was necessary to ensure that the agreement would not interfere with their ability to carry out their responsibilities for prudential regulation and supervision. Financial regulators also insisted that any dispute settlement panel dealing with financial services must have the appropriate expertise regarding the specific financial service at issue. Ultimately, provisions addressing both of these issues were included in the GATS Annex on Financial Services.

The prudential carve-out allows a country to take prudential measures "to ensure the integrity and stability of the financial system" or "for the protection of investors, depositors, policy holders or persons to whom a fiduciary duty is owed" even if the measures are inconsistent with other provisions of the GATS. The carve-out also contains an anti-abuse provision. As a result, not only the question of whether a particular measure is prudential in the first place but also the question of whether a seemingly "prudential" measure is being misused to avoid a country's obligations and commitments could be brought before a WTO dispute settlement panel. However, as of October 2007, there had been no dispute settlement proceeding and no request for consultation on a financial services issue. Therefore the scope of the prudential carve-out and its antiabuse provision remain untested in WTO jurisprudence.

A GATS commitment is permanent in the sense that it cannot be withdrawn without compensation of trading partners. Failure to honor a GATS commitment could open a country to a WTO dispute settlement proceeding and, ultimately, to WTO-sanctioned retaliatory measures by its trading partners. Thus backsliding in the face of domestic political pressures could be extremely costly. The GATS calls for periodic negotiating rounds to improve commitments for financial and other services to achieve "a progressively higher level of liberalization." The Doha Round, which was launched in November 2001 but as of October 2007 remained stalled, includes a new set of negotiations on trade in services.

Ideally, GATS negotiations can both reinforce and build on market and political forces that are creating pressures for liberalization within a host country. In this regard, a country's "readiness" for reform is critical. Indeed, undertaking commitments in the GATS can be an integral part of a country's longer-term policy reform agenda. For example, China's WTO accession commitment to open its banking market to FDI within five years (i.e., by December 11, 2006) in effect set a domestic political deadline for substantial progress in reforming China's banking system. During the 1997–98 Asian financial crisis, making commitments in the GATS to permit FDI in their domestic financial sectors offered an additional opportunity for Asian governments to try to reassure markets that they were committed to longer-term policy reform.

In the 1997 agreement on financial services commitments, most of the WTO members participating in that agreement "bound" the levels of liberalization for FDI that existed as the negotiations entered their final phase in late 1997. For a number of emerging market economies and other developing countries, such liberalization represented a substantial improvement over the liberalization that existed a few years earlier. Moreover, capturing existing levels of liberalization in binding commitments subject to WTO dispute settlement is important in its own right. In contrast to the commitments for FDI, the 1997 commitments for cross-border services provided "from" the country of the service supplier ("cross-border supply") were, even for OECD countries, relatively limited and did not always reflect existing liberalization. Most countries that have acceded to the WTO since its inception in 1995 have made strong financial services commitments, including new liberalization, for FDI and, in some cases, cross-border services.

Three Pillars of Liberalization The inclusion of financial services in the GATS and the negotiation of the 1997 agreement on financial services commitments constitute a significant step in the larger process of achieving international contestability of markets for financial services and strengthening domestic financial systems. "International contest-

ability of markets" refers to the creation of markets that are competitive and efficient on a global basis a goal that can be achieved by removing all types of barriers to foreign participation in host-country markets. International contestability is, in effect, based on three pillars of liberalization namely, (1) opening markets by ensuring "national treatment" and "market access"; (2) removing nondiscriminatory structural barriers (domestic structural reform); and (3) liberalization of capital movements.

For financial services, the GATS has so far dealt mainly with first-pillar liberalization, that is, liberalization aimed at opening markets to foreign services and service suppliers and ensuring that they enjoy substantially the same treatment as their domestic counterparts. First-pillar barriers of particular importance for financial services include limitations on foreign ownership interests in domestic financial firms; restrictions on juridical form, for example, prohibiting a foreign financial firm from establishing a branch as opposed to a subsidiary; and prohibiting branches or subsidiaries of foreign financial firms from engaging in all of the activities that are permissible for their domestic counterparts.

One question for future negotiations on financial services is whether, or to what extent, they should extend into the second pillar. Second-pillar liberalization deals with nonquantiative structural barriers that do not discriminate between domestic and foreign services and service suppliers. In contrast to national treatment and market access which ensure that foreign services and services suppliers can enter a host-country market as currently structured and enjoy equality of competitive opportunities vis-à-vis their domestic counterparts second-pillar liberalization represents an effort to create maximum potential competitive opportunities in a host-country market. It may therefore require major domestic structural reform. Second-pillar barriers can be created by anticompetitive domestic policy measures, the inadequacy or absence of domestic regulation, differences in national rules that make it difficult to conduct operations on a global basis, or lack of regulatory transparency. Moreover, second-pillar barriers to trade in financial services are not limited

to financial sector regulation. They also include, for example, a host country's lack of an effective legal framework for corporate governance or insolvency; both of these areas are part of the international work on strengthening domestic financial systems.

The European Union's single-market program represents the most far-reaching effort to date to remove nondiscriminatory structural barriers among a group of nations. Predicated on political agreement on goals for economic liberalization, that effort is being carried out in the context of the unique supranational legislative, judicial, and administrative structure of the European Community. Even within the European Union, however, important nondiscriminatory structural barriers to trade in financial services among the member states still exist, especially with regard to retail financial services.

At present, the GATS addresses certain types of second-pillar barriers, including lack of transparency and barriers created by domestic regulation, in very general terms. Some second-pillar barriers are also dealt with in countries' schedules of commitments. For financial services, most of these commitments are simply "best efforts" commitments. Many believe that it is not realistic or appropriate to negotiate and bind in the GATS additional second-pillar liberalization for financial services. However, one type of second-pillar liberalization that involves procedural as opposed to substantive barriers namely, regulatory transparency is being discussed in the Doha Round for possible inclusion in the GATS. Disciplines on regulatory transparency, which might be designed to apply to all service sectors or only to financial services, would strengthen current GATS disciplines on transparency for example, by requiring WTO members to establish a meaningful procedure for interested parties to comment on a proposed regulation prior to its adoption in final form. Such a requirement, together with other transparency disciplines, was included in China's Protocol of Accession to the WTO; similar disciplines have been included in the financial services chapters of bilateral free trade agreements to which the United States is a party.

The GATS deals with third-pillar liberalization only insofar as it affects countries' specific commitments to liberalize trade in services; in general, liberalization of capital movements is a matter of concern for the IMF. In particular, the GATS prohibits a WTO member from imposing restrictions on certain capital movements that would be inconsistent with its specific commitments to provide national treatment and market access. As a result, if a country makes a commitment to liberalize trade with respect to a particular financial service in the GATS, it is also making a commitment to liberalize most capital movements associated with the trade liberalization commitment. The country is not, however, making an across-the-board commitment to freedom of capital movements. The GATS provisions dealing with capital movements, like GATS-specific commitments to liberalize trade in services, are subject to a balance-of-payments safeguard. Both the capital movements and balance-of-payments safeguard provisions of the GATS refer to and are consistent with the IMF's responsibilities in these areas.

Trade Liberalization and Sound Financial Systems The financial services sector including international trade in financial services plays a vital role in the modern world economy. The adoption of international minimum standards and codes for sound financial systems and their implementation by individual countries make an important contribution to financial stability, both nationally and globally. These standards and codes also provide a strong foundation for liberalization of trade in financial services in the GATS and in regional and bilateral trade agreements. As the GATS explicitly recognizes, liberalization of trade in financial and other services is an ongoing process. For financial services, this process is being driven in large part by market forces and new technologies. It also reflects a growing recognition among policymakers that opening markets to foreign suppliers of financial services can both benefit host-country consumers and contribute to the resiliency of domestic financial systems.

See also capital controls; General Agreement on Trade in Services (GATS); International Monetary Fund (IMF); sequencing of financial sector reform; trade in services; World Trade Organization

FURTHER READING

Committee on Global Financial Systems. 2004. *Foreign Direct Investment in the Financial Sector of Emerging Market Economies*. Report of Working Group. Basel: Bank for International Settlements. Downloadable from http://www.bis.org/publ/cgfs22.pdf?noframes=1. Explores issues that financial sector FDI raises for investing institutions, home and host countries, and the global financial system.

Dobson, Wendy, and Pierre Jacquet. 1998. *Financial Services Liberalization in the WTO*. Washington, DC: Institute for International Economics. Examines benefits of financial reform, including pitfalls and transitional issues, and analyzes 1997 agreement on GATS financial services commitments.

Financial Stability Forum. 2002. "12 Key Standards for Sound Financial Systems." Downloadable from http://www.fsforum.org/compendium/key standards for sound financial system.html. Twelve international minimum standards and codes of good practices are identified by the Financial Stability Forum (FSF) as key for sound financial systems and deserving of priority implementation, depending on country circumstances.

Graham, Edward M., and Robert Z. Lawrence. 1996. "Measuring the International Contestability of Markets: A Conceptual Approach." *Journal of World Trade* 30 (5): 5 20. Analytic framework for the concept of international contestability of markets.

International Monetary Fund, Independent Evaluation Office. 2006. *Report on the Evaluation of the Financial Sector Assessment Program*. Washington, DC: International Monetary Fund. Downloadable from http://www.imf.org/External/NP/ieo/2006/fsap/eng/index.htm. Analyzes effectiveness of Financial Sector Assessment Program (FSAP) from the perspective of the IMF.

Kampf, Roger. 1998. "Financial Services in the WTO: Third Time Lucky." *International Trade Law and Regulation* 4 (3): 111 123. Analyzes 1997 agreement on GATS financial services commitments.

Key, Sydney J. 1997. *Financial Services in the Uruguay Round and the WTO*. Occasional Papers 54. Washington, DC:

Group of Thirty. Analyzes issues and challenges in the Uruguay Round financial services negotiations.

———. 2003. *The Doha Round and Financial Services Negotiations.* Washington, DC: AEI Press. Down loadable from http://www.aei.org/book456. Analyzes the role of the GATS and the WTO in liberalization and regulation of the financial services sector from both trade and financial regulatory policy perspectives and identi fies six broad goals for the Doha Round financial services negotiations. Current entry is adapted from this book.

Kono, Masamichi, Patrick Low, Mukela Luanga, Aaditya Mattoo, Maika Oshikawa, and Ludger Schuknecht. 1997. *Opening Markets in Financial Services and the Role of the GATS.* Geneva: World Trade Organization. Downloadable from http://www.wto.org/english/res e/ booksp e/special study 1 e.pdf. Examines benefits and challenges of liberalizing trade in financial services and the role of the GATS.

Mattoo, Aaditya. 2000. "Financial Services and the WTO: Liberalization Commitments of Developing and Tran sition Economies." *The World Economy* 23 (3): 351 86. Studies the financial services commitments made by nearly 100 developing countries and all Eastern Euro pean members of the WTO.

Rogoff, Kenneth S. 2002. "Rethinking Capital Controls: When Should We Keep an Open Mind?" *Finance and Development* 39 (4): 55 56. Downloadable from http:// www.imf.org/external/pubs/ft/fandd/2002/12/rogoff .htm. Perspective on capital controls from former chief economist of the IMF.

Sato, Setsuya, and John Hawkins. 2001. "Electronic Fi nance: An Overview of the Issues." In *Electronic Finance: A New Perspective and Challenges.* BIS Papers 7. Basel: Bank for International Settlements. Downloadable from http://www.bis.org/publ/bispap07a.pdf. Overview of issues based on presentations and discussions at work shop on electronic finance organized by the Bank for International Settlements.

Skipper, Harold D., Jr. 2001. *Insurance in the General Agreement on Trade in Services.* Washington, DC: AEI Press. Overview of insurance and insurance markets, international trade in insurance services, and insurance commitments in the GATS.

SYDNEY J. KEY

■ **first-mover advantage**

See foreign market entry

■ **fixed costs and foreign direct investment**

Fixed costs are production costs, such as research and development (R&D) expenditures and rent of pro duction facilities, that do not vary by the quantity of output. The existence of fixed costs often suggests increasing returns to scale in the production process: the average fixed cost borne by each unit of output decreases as output expands.

Fixed costs are a key element in a firm's decision to invest abroad and set up foreign subsidiaries. How motivated firms are to invest in a foreign country can be altered by the existence of fixed costs. Firms may decide to establish plants in foreign countries to avoid trade costs or to take advantage of lower factor prices. The former strategy leads firms to duplicate production across countries and undertake a horizontal type of foreign direct in vestment (FDI), while the latter motivates firms to locate different stages of production in different countries and engage in a vertical type of FDI. When fixed costs arise at the level of a plant, they create a disincentive to have duplicated plants, and thus a disincentive for horizontal FDI.

Firm-Level versus Plant-Level Fixed Costs Fixed costs may exist at the firm and/or the plant level. This distinction is particularly important to multi-plant firms. Plant-level fixed costs include any fixed costs that accrue based on the number of plants, such as the cost of renting production facilities and the cost of equipment. These fixed costs provide physical capital for a specific plant and increase with the number of plants. By implication, they discourage firms from maintaining multiple plants and encour age geographic concentration of production.

Firm-level fixed costs, in contrast, include any fixed costs that accrue based on the number of firms, such as R&D and advertising costs. This type of fixed cost does not increase with the number of plants and thus raises firms' incentive to expand horizontally and spread the cost across multiple plants.

More specifically, the role of fixed costs in a firm's decision to invest abroad depends on whether the costs involve the supply of knowledge capital or physical capital. Knowledge capital such as technology know-how can be simultaneously supplied to multiple production locations and thus does not have to increase with the number of plants a firm operates. Physical capital such as production facility, however, is exclusively supplied to one location and thus requires additional investments when a firm establishes new plants. FDI is more likely to arise in industries where knowledge-based assets are important and scale economies in physical capital are less critical that is, where firm-level fixed costs are important relative to plant-level fixed costs. In industries that are both knowledge- and physical-capital intensive, firms invest abroad when the benefit of proximity to consumers or lower factor prices in the host country dominate the plant-level diseconomies of scale.

Theoretical Considerations Numerous theoretical studies, led by the seminal work of Markusen (1984) and Horstmann and Markusen (1987, 1992), have investigated the role of these two types of fixed costs in a firms' decision to undertake FDI. Markusen (1984) was one of the first authors who showed from an international trade perspective that multinational enterprises may arise endogenously in equilibrium. In his model, firms may separate headquarter activities such as R&D from production. The headquarter activities require an up-front fixed cost and in turn provide knowledge-based assets, such as patents, blueprints, and trademarks. These assets then serve as a joint input that can be supplied to all the production facilities of a firm without lowering the value of the input. This approach is extended by Horstmann and Markusen (1987, 1992), who show, for instance, that two-way investment flows between pairs of countries can arise in equilibrium. These studies find that firm-level scale economy implies that it is best to have one firm produce for all markets as doing so avoids the need for duplication in headquarter services. A multiplant multinational firm with plants in both countries or a firm with a plant in one country that exports to the other would have cost efficiency over a single-plant firm in each country selling locally. Furthermore, firms are motivated to set up local production plants in foreign countries rather than export from headquarters when the trade costs are high or the variable production cost in the foreign country is low.

The existing theoretical literature also shows that plant-level scale economies give rise to export-platform FDI. Ekholm, Forslid, and Markusen (2007), for example, find in both their model and empirical results that, in the presence of large plant-specific fixed costs, firms tend to adopt an export-platform strategy in which they concentrate their production in a low-cost country and export to other markets. A reduction in trade costs between the low-cost country and other countries would further augment firms' incentive to undertake geographically centralized FDI and achieve economies of scale.

Empirical Assessment The effect of fixed costs on multinational firms' decision to invest abroad has also been investigated empirically. In particular, Brainard (1997) and Yeaple (2003) examine the extent to which multinational production location decisions can be explained by a trade-off between maximizing proximity to consumers (or exploiting countries' difference in comparative advantage) and concentrating production to achieve economies of scale. Their analyses consider both firm-level and plant-level economies of scale. Firm-level scale economies are measured by the number of non-production workers in the average U.S.-based firm in each industry, while plant-level scale economies in each industry are measured by the number of production workers in the average U.S. plant. As predicted by the theory, these studies find that overseas production by multinational firms increases relative to exports when the scale economies at the plant level are lower than those at the firm level.

Ekholm, Forslid, and Markusen (2007) also confirm the role of scale economies in multinationals' investment decisions. They examine specifically export-platform FDI, a product of significant plant-level fixed costs and declining trade costs. They find that since the removal of trade barriers within the European Union and the North American Free Trade Area, multinational firms are increasingly

motivated to concentrate their production in a low-cost member of the integrated region and supply the rest of the region through exports.

In sum, theories of the multinational firm suggest that firm-level fixed costs give rise to increased FDI, while plant-level fixed costs encourage export-platform FDI but deter horizontal FDI. Broad empirical support exists for these predictions.

See also footloose production; foreign direct investment and exit of local firms; foreign direct investment under monopolistic competition; foreign direct investment under oligopoly; foreign market entry; knowledge-capital model of the multinational enterprise; proximity-concentration hypothesis; subsidies and financial incentives to foreign direct investment

FURTHER READING

Brainard, S. Lael. 1997. "An Empirical Assessment of Proximity Concentration Trade off between Multi national Sales and Trade." *American Economic Review* 87 (4): 520 44. One of the earliest empirical studies that examine the role of trade costs and scale economies in multinationals' investment decisions.

Ekholm, Karolina, Rikard Forslid, and James Markusen. 2007. "Export Platform Foreign Direct Investment." *Journal of European Economic Association* 5 (4): 776 95. An important study of the causes of export platform FDI and the effect of free trade areas on multinationals' location decisions.

Horstmann, Ignatius, and James Markusen. 1987. "Stra tegic Investments and the Development of Multi nationals." *International Economic Review* 28 (1): 109 21. A first theoretical analysis that investigates the suf ficient production and transportation conditions for multinational corporations to arise.

——. 1992. "Endogenous Market Structures in Inter national Trade." *Journal of International Economics* 32 (1 2): 109 29. A systematic investigation of market structure and the existence of multinational and ex porting firms.

Markusen, James. 1984. "Multinationals, Multiplant Eco nomics, and the Gains from Trade." *Journal of Inter national Economics* 16 (3 4): 205 26. One of the first theoretical contributions on the efficiency and welfare implications of multiplant multinational firms.

Yeaple, Stephen. 2003. "The Role of Skill Endowments in the Structure of U.S. Outward Foreign Direct Invest ment." *Review of Economics and Statistics* 85 (3): 726 34. A careful account of the industry and country determi nants of U.S. outward FDI with an emphasis on the role of skilled labor endowment.

MAGGIE XIAOYANG CHEN

■ **floating exchange rate**
See exchange rate regimes

■ **footloose production**
Multinational firms are increasingly important in the world economy. Since multinationals are usually firms with plants in more than one country they are considered "footloose" since they can shift their ac tivities more easily from one country to another. At the end of 1997 the added value of all multinational enterprises (MNEs) in the world was approximately a quarter of the total added value in the world. In 2002, the stock of foreign investments worldwide was on average 22 percent of the gross domestic product of each country, and the corresponding percentages for West European countries were still higher, between 31.4 and 42.7 percent.

Increasingly, worldwide, more and more trade is in the hands of MNEs. The amount of intrafirm trade within MNEs accounts for almost one third of all worldwide trade. The number of MNEs also has increased substantially. Since the mid-1990s the amount of worldwide foreign direct investment has increased more rapidly than exports.

By the very act of becoming a multinational, the production of MNEs is potentially footloose in na ture. A distinction can be made between horizontal MNEs, which produce the same goods and services in multiple countries, and vertical multinationals, which fragment production geographically by stages. While theoretically this distinction is easy to make, it is much harder to analyze empirically how much horizontal versus vertical MNE activity is involved

(Carr et al. 2001). Nonetheless, both types of multinational activity enable MNEs to shift production across countries.

MNEs are flexible in terms of geographically organizing their production structure. Optimizing their activities over different countries often results in vertical fragmentation of production, with activities of R&D, design, marketing, and distribution usually located in high-income developed countries and the production of intermediate inputs and assembly located in low-wage/low-tax countries (Hanson et al. 2005). This fragmentation of production results from endowment differences and falling transportation, tariff, and information costs across countries. In response to changes in any of these causes, MNEs may respond by shifting more (or less) production abroad. For example, improvements in information technology or increases in domestic wages may cause an MNE to shift the production of intermediate components abroad.

MNEs also often establish plants undertaking similar activities in multiple countries. A typical MNE may have several similar plants, with one in each major market, such as one in Asia, one in North America, and one in Europe. This duplication of plants is generally undertaken to avoid transportation costs, tariffs, and other trade restrictions. Although these plants are intended primarily to serve the immediate market (the country where located and neighboring countries), they can also be used to supply other markets if needed. A shock in demand or supply in one region could be offset by adjustments in production throughout the global network of plants.

How Footloose Is Production? Although all MNEs have the potential to shift production across countries, how easy it is to do so can differ. The ownership share that MNEs have in affiliates with plants abroad varies. While some studies consider ownership of 20 percent or more, other studies consider a higher cutoff for MNE ownership in the affiliate. For an MNE to be really footloose, however, arguably ownership in the affiliate by the MNE should exceed 50 percent or more of the shares. Only when the MNE has majority ownership in its affiliate

can it easily decide to "pack up and go" as typical for footloose production. Also, the higher the number of majority-owned affiliates, the more footloose the MNE is since it can relocate production activities from one affiliate to another.

There are two main reasons that production is becoming increasingly footloose. First, the production of MNEs is more footloose than that of domestic firms, so as the degree of multinational activity increases, production overall becomes more able to be moved across national borders. Second, changes may be occurring that make production by MNEs more flexible. New markets are joining the world economy and opening up to the inflow of foreign capital. This growth in market opportunities implies that MNEs geographically expand the scope of their networks with more plants in more countries.

Implications of Footloose Production The footloose nature of multinational production has implications for local product, labor, and capital markets. Whatever the type of MNE involved, its presence in the host country tends to increase the volatility of the local economy. MNEs' production tends to respond more quickly to upswings or downswings in economic conditions than does the production of domestic firms. Due to their footloose nature, MNEs are usually the first firms to come when conditions are favorable and the first to exit the market when fortune reverses. The increase in the number of MNEs worldwide therefore makes local economies more subject to shocks and volatility. Countries with flexible labor markets, that is, with low hiring and firing costs, are better equipped to cope with this new challenge since they can more quickly and at lower costs adjust to shocks.

In terms of local labor markets, MNEs usually have a stronger bargaining position with labor unions than domestic firms do, since MNEs always have the outside option of relocation. This does not imply that MNEs necessarily pay lower wages, however. Empirical evidence seems to suggest that MNEs pay higher wages and provide better working conditions than local firms in developing countries (Brown et al. 2003). In developed countries, wages in MNEs also tend to be higher. One reason is that MNEs tend to be

larger and more productive; they also produce higher-quality goods than their local rivals, which gives them greater market power in product markets and allows them to charge higher prices. Higher product market rents are likely to result in higher wages. MNEs on average employ better-skilled workers to guarantee higher quality in production and a higher level of productivity. They often invest substantially in internal training. All these elements offer explanations for the wage gap between MNEs and local firms in the same country or region.

Theory suggests that MNEs in high-income countries export unskilled jobs to their affiliates in low-wage countries. There is some empirical confirmation; however, Braconier and Ekholm (2000) and Konings and Murphy (2006) find evidence of substitutability of labor among high-wage countries, but not between high-wage and low-wage countries. Employment effects may depend on the type of activity that is outsourced. Some suggest that in the long run the outsourcing of unskilled activities could create more job opportunities in developed countries. This subject is hotly debated, but more research is needed to understand the dynamic effects of MNEs on the labor markets of parent and host countries.

MNEs rely less on local capital markets. Typically MNEs and their affiliates benefit from internal capital markets. Recent evidence suggests that MNEs are less subject to credit constraints than other firms (Harrison et al. 2004). Hence MNEs suffer less from incomplete capital markets in developing countries than local firms since they can always turn to intrafirm financing. They depend less on local banking conditions and on the competition in the banking sector.

The footloose nature of multinational production implies that MNEs usually also have a strong bargaining position vis-à-vis governments that host their affiliates. Prior to choosing where to locate a plant, MNEs are often able to negotiate financial incentives with host governments. In recent years national governments have increasingly been using the corporate tax instrument as a tool to attract MNEs into their region. Small countries and countries that are

otherwise less attractive may have to offer more favorable tax treatment to MNEs than to local firms. Vandenbussche and Tan (2005) found that in a small country such as Belgium, foreign-owned firms on average have lower effective tax rates than purely domestic firms.

In sum, the increasing prevalence of MNEs has caused production to become increasingly footloose in nature, with important implications for the world economy.

See also globalization; multinational enterprises

FURTHER READING

Braconier, Henrik, and Karolina Ekholm. 2000. "Swedish Multinationals and Competition from High and Low Wage Countries." *Review of International Economics* 8 (3): 448 61. Examines whether there is substitution between employment in different parts of Swedish firms.

Brown, Drusilla K., Alan V. Deardorff, and Robert M. Stern. 2003. "The Effects of Multinational Production on Wages and Working Conditions in Developing Coun tries." NBER Working Paper No. 9669. Cambridge, MA: National Bureau of Economic Research. Assesses evidence regarding the effect of multinational production and working conditions in developing countries.

Carr, David L., James R. Markusen, and Keith E. Maskus. 2001. "Estimating the Knowledge Capital Model of the Multinational Enterprise." *American Economic Review* 91 (3): 693 708. Tests the horizontal versus the vertical multinational model.

Hanson, Gordon H., Raymond J. Mataloni, and Matthew J. Slaughter. 2005. "Vertical Production Networks in Multinational Firms." *Review of Economics and Statistics* 87 (4): 664 78. Studies intrafirm trade in intermediate inputs.

Harrison, Ann E., Inessa Love, and Margaret McMillan. 2004. "Global Capital Flows and Financing Con straints." *Journal of Development Economics* 75 (1): 269 301. Examines in a cross section of countries the relation between inward foreign direct investment and domestic firms' financing constraints.

Konings, Jozef, and Alan Murphy. 2006. "Do Multi national Enterprises Relocate Employment to Low wage Regions? Evidence from European Multinationals." *Review of World Economics* 142 (2): 267 86. Empirical

analysis of employment substitution between parent firms and foreign affiliates.

Vandenbussche, Hylke, and Chang Tan. 2005. "The Taxation of Multinationals: Firm Level Evidence for Belgium." LICOS discussion Paper No. 160/2005, Catholic University Leuven. http://www.econ.kuleuven .be/LICOS/DP/DP2005/DP160.pdf. An empirical assessment of whether multinational enterprises are given more favorable tax treatment compared to domestic firms.

HYLKE VANDENBUSSCHE

■ foreign direct investment (FDI)

Foreign direct investments are defined as investments in which a firm acquires a majority or at very least a controlling interest in a foreign firm. Foreign investments not involving a majority or controlling stake are typically referred to as portfolio investments. Firms making foreign direct investments (FDI) are referred to as multinational enterprises (MNE) and the two terms are used somewhat interchangeably. A direct investment may involve creating a new foreign enterprise, often referred to as a greenfield investment, or acquiring an existing foreign firm (sometimes referred to as a brownfield investment, though that term is much less common; *acquisition* is the typical label).

Historically, there are three strands of literature that see the multinational and FDI in different ways: the international business tradition, the trade-theory tradition, and the macroeconomic tradition. This entry will focus heavily on the trade-theory tradition, where the biggest developments in the last twenty years have occurred.

The international business approach is very individual-firm oriented. It details the determinants of the decision of firms to go abroad and the mode they chose for doing so. In addition to FDI, the firm considers exporting, joint ventures, licensing or contracting with arm's-length foreign firms, and so forth. The international business literature has been far more interested than the other streams of literature in the choice-of-mode decision.

It is probably accurate to say that until the late 1980s the microeconomic trade-theory approach to FDI and the macroeconomic tradition were pretty much the same. These two traditions did not really distinguish between direct and portfolio investments: there was no real attempt to model the "D" in FDI. Both schools modeled FDI as the movement of homogeneous capital from locations where its return was relatively low to where its return is higher. The simple approach to capital flows had a natural intersection with Heckscher-Ohlin trade theory, in which factors are expensive where they are scarce and cheap where they are abundant. The consequence is the obvious hypothesis that capital should flow from capital-rich to capital-scarce countries.

There was no sense of individual firms in this literature, and certainly no modeling of mode choice. Even trade theory, with its better-developed sense of general-equilibrium than macroeconomics, was dominated by perfect-competition, constant-returns-to-scale models in which individual firms had no real meaning. But trade theory did have advantages over the international business approach in that it had a basic general-equilibrium structure that did at least give some predictions as to the pattern of capital flows we should observe.

Macroeconomics has more or less continued in the tradition of restricting analysis to aggregate capital flows generated by international rental-rate or cost-of-capital differentials. It is not easy to fit a rich structure for individual firms into macro models, and hence that stream of literature continues to make no real distinction between FDI and portfolio investments.

International trade theory, on the other hand, began to move sharply away from the macro approach in the 1980s, and to draw a clear distinction between FDI and portfolio investments. It began to move more toward the international business literature in that it included meaningful treatments of individual firms, yet the trade approach retained the general-equilibrium roots of its tradition. The split with macro seems to have been driven by some important statistical evidence that casts considerable doubt on the suitability of cross-country differences

in the cost of capital as a driving and motivating force for FDI.

Troubling Statistics The first statistical difficulty confronting the traditional theory is that the high-income developed countries are not only the major source of FDI but also the major recipients. FDI does not primarily flow from capital-rich to capital-poor countries; it flows primarily from capital-rich to other capital-rich countries. Firms from high-income countries are mutually invading one another's markets. In addition, the FDI that does flow to developing countries is highly concentrated in the most advanced of those countries.

Many statistics are found in Caves (2007), Markusen (2002), UNCTAD, OECD, BEA, and other publications. One simple way to measure this source/recipient pattern is to simply compute the shares of total world FDI inward and outward stocks that are found in different countries or groups of countries. UNCTAD statistics for 2003 indicate that the developed countries accounted for 89 percent of the outward stock in this year, but also accounted for 69 percent of the inward stock. Of course, the developed countries also account for the overwhelming share of world income, so we can instead divide the inward share of a group by its share of total world income. This yields the results shown in table 1. The high-income countries are close to a share of inward FDI equal to their share of income. Developing countries are higher on this score, as the simple capital-scarcity macro approach would suggest, but it turns out that this is highly concentrated in the more advanced of the developing countries. The least-developed countries (a UN-defined group of 44 countries) have a share of inward FDI only slightly greater than their share of world income.

A second statistic that led to a rethinking of trade theory is that there is very often a disconnect between the act and amount of FDI and the sources of financing for new investments or acquisitions. There is much less systematic evidence on this point since published data rarely comment on the source of financing for individual foreign projects. Yet the international business literature documents a great variety of financing choices: retained earnings of the parent firm, equity or bonds issued in the parent-country market, equity or debt issued in the host-country market, or third-country financing. Apparently, many FDI projects in China are financed with Chinese capital, although systematic data are not known to this author. If we think of Chinese debt financing an American direct investment, we have in fact a portfolio and a direct investment flowing in opposite directions.

A third relevant statistic is that FDI is attracted to large markets. Inward FDI divided by GDP should not display any particular pattern in a simple cost-of-capital macro approach. The data shown in table 2 are from 1993 and thus somewhat dated (Markusen 2002), but there is no reason to think that the basic message has changed. They show that not only is FDI attracted to rich countries, it is clearly attracted to large markets.

A fourth statistic that called for a new approach to FDI is that a very large proportion of the actual output of FDI projects is for local sale in the host country. Firms were often replicating their home

Table 1
Share of world inward FDI stock divided by share of total world income

Developed countries	0.90
Developing countries	1.37
Least developed countries	1.07

Table 2
Inward FDI per capita

Country group GDP per capita (US$)	Country group Size	Average inward FDI per capita (US$)
> 5000	Large	242
	Small	54
2500 5000	Large	46
	Small	32
1200 2500	Large	33
	Small	31
600 1200	Large	11
	Small	3

Table 3
Sales of foreign manufacturing affiliates
of U.S. multinationals, 2000 (shares in total sales,
sample of 39 countries for which BEA data is available)

	Local sales	Export sales to the U.S.	Export sales to to third countries
all countries	0.60	0.12	0.28
Ireland	0.13	0.16	0.71
Canada	0.57	0.38	0.05
Mexico	0.53	0.39	0.08

activities, goods or services, in foreign markets. These are now generally known as "horizontal" investments, although the term "market-seeking" is also used. The latter term emphasizes that the FDI is not motivated by cost-side considerations as it would be in a cost-of-capital-differential approach, but rather by the demand-side motive of serving the local market. Table 3 presents a few statistics from BEA data. Perhaps 60 percent local sales does not seem terribly high for the sample as a whole. But more interesting is the fact that only 12 percent of sales are back to the United States; 28 percent of the sales are to third countries, often referred to as "export-platform" production and sales.

Closer inspection of the data suggests that third-country sales are most important for affiliates located in large regionally integrated trade areas: the exports are destined for the other regional members. In an important sense, this represents horizontal production in which one location within the region is chosen. United States affiliates in Ireland, for example, export 71 percent of their total production to third countries, and apparently almost all of this is horizontal production destined for other European Union markets. For U.S. affiliates in Canada and Mexico, exports are similarly directed to other members of the North American Free Trade Agreement, and so the overwhelming portion of their exports are to the United States. Thus is it clear that the overwhelming portion of affiliate output is destined for local or regional sale. Note that Canada and Mexico

exhibit remarkably similar statistics in spite of the large difference in per capita income, and that local sales have shares close to the world average for U.S. affiliates. While many investments in Mexico are surely chosen to access cheap labor for production for the United States market, the general notion that multinationals are firms seeking cheap production abroad for sale back home is at best a minor phenomenon.

The consequence of these observations was that trade theory needed to move away from the old macro approach and adopt a firm-based approach, yet one still rooted in the general-equilibrium tradition. Trade theory began to think of and indeed measure FDI not in terms of the value of investments (inputs), but in terms of the outputs of foreign affiliates, the destinations for those outputs, and the trade patterns with the parent firm. But the old bias is still very evident in data sources: it is much easier to get data on FDI stocks and flows than on affiliate outputs, sales destinations, and intrafirm trade.

The International Business Approach As suggested earlier, researchers in the field of international business have produced a rich set of theory and empirical analysis, though little is formalized and few testable hypotheses emerge. Nevertheless, the newer formal models owe a great debt to these scholars. An early approach was by Dunning (1973) with his ownership-location-internalization (OLI) framework. Dunning suggested that three conditions must be met before a firm will want to establish an owned production facility. Subsequent researchers have assembled a great body of evidence, though very little of it formal econometric work, about the form that these advantages take.

The first condition is ownership advantage. Given a disadvantage relative to local firms in a host country, a foreign firm must own a propriety asset such as a superior product, production process, patent, trademark, or asset that gives it a compensating advantage over local firms in the host market. This focuses the theory on the assets of the individual firm and away from some general return to homogeneous capital. Caves (2007) has written much on this idea, using the term *intangible assets* to label these proprietary assets.

Empirical analysis established that multinationals tended to be firms that are intensive in knowledge-based assets (Markusen 2002) rather than physical capital. Multinationals are associated with patents, research and development (R&D) intensity, skilled white-collar and technical workers and engineers, new and complex products, and product differentiation variables such as advertising, trademarks, and brand names. Multinationals have a high value of intangible assets, which can be measured as a sort of Tobin's Q: the ratio of the market value of the firm to the book value of capital. There are good reasons why multinationality should be associated with knowledge-based assets, but a discussion of this is temporarily postponed.

But ownership advantage is not sufficient. For example, if the purpose of the investment is to serve local markets, then the firm can exploit its asset through exporting. The second condition is therefore location advantage. This is some factor that leads the firm to prefer to actually produce abroad rather than export. Location advantages tend to depend in large part on whether the purpose of the investment is to serve local markets or to export from the host country. For the first type, termed horizontal or market-seeking investments as noted earlier, location advantages are (1) a large host-country market to compensate for set-up costs and plant-level scale economies, and (2) trade costs in the form of tariffs or transportation costs (time as well as money) that make serving the host country by exports expensive. For investments that are more directed at using the host country as an export platform, termed vertical or resource-seeking investments, location advantages are more in the form of low input costs and low trade costs to get intermediate and final goods into and out of the country.

The third condition is internalization advantage. The firm must have a reason to own the foreign production facility rather than simply to license its asset or contract with a local firm to produce on its behalf. Internalization is often contrasted with its mirror image, outsourcing. These are the two alternatives to one decision: the firm must choose between internalizing and outsourcing.

Internalization advantages are the most abstract of the three. For some authors, the principal issue derives from properties of knowledge capital, which are discussed a bit more below. The firm needs to maintain tight control over knowledge-based capital or the value can be easily dissipated through copying and other forms of agent opportunism. Many threats of asset dissipation arise from the lack of strong legal institutions in host countries, such as intellectual property protection and contract enforcement. Other determinants of the internalization/outsourcing decision are familiar from more general discussion of the boundaries of the firm and are not focused on anything particularly international in scope (i.e., strictly domestic firms face decisions on what activities to outsource and which to internalize).

In what follows, we will focus on ownership and location issues in discussing both theory and empirical evidence. The goal is to explain the statistics presented earlier. Internalization is attacked with a very different set of tools and, to date, empirical evidence is scarce.

Early Trade-Theory Models Two very different papers appeared in 1984. Helpman (1984) is a model in which a firm can decompose a production process into a headquarters activity and a production activity. Headquarters and production have different factor intensities and firms can choose to geographically separate these activities. A multinational is a firm that has its headquarters in one country and a plant in the other; in other words, this is a model of vertical multinationals. There are no trade costs in the model and multinationals arise only when countries are sufficiently different in relative factor endowments (i.e., countries must be outside the factor-price-equalization set in the Edgeworth box). In particular, in this model multinationals cannot arise between identical countries, and there are no multiplant firms. Helpman's model is very much in the older tradition of factor-price difference driving FDI discussed earlier.

Markusen (1984) is a model in which there is also something like a headquarters activity and a production activity that can be geographically separated, but these activities do not have different factor intensities.

The whole focus of this model is different from Helpman's. The key idea in the Markusen paper is that headquarters activity, such as R&D, has a "jointness" or "public-goods" property, in that it can yield the full value of its productivity in two locations: adding a second production facility does not reduce the value of the R&D asset in the first location. A blueprint, formula, or procedure can be jointly and fully used in multiple locations. The focus of Markusen's paper is on multinational, horizontal or market-seeking firms that produce in two locations to exploit the value of their knowledge capital. Multinationals can arise between two identical countries, marking a clear break with the old capital-flows literature.

Subsequent Theoretical Developments Important refinements of the horizontal approach are in Horstmann and Markusen (1992) and Brainard (1993). These papers solve for equilibrium market structure between two countries and show that two-plant horizontal multinationals arise when firm-specific fixed costs (knowledge capital) are important relative to plant-level fixed costs, when trade costs are high, and when the two markets are large and similar in size. The result that FDI is expected between large, similar countries provided a theoretical underpinning for the empirical finding that most FDI occurs between large, high-income countries as discussed earlier.

The general-equilibrium structure of the problem was later fleshed out in several papers by Markusen and Venables (1998, 2000). Markusen and Venables use the world Edgeworth box. They concentrate on horizontal two-plant firms and do not consider vertical structures in which a single plant and headquarters are located in different countries. There can be single-plant national firms and two-plant multinationals located in each country, or four potential firm types in all.

They solve the model over a grid in the Edgeworth box, where countries different in size and/or in relative endowments occur at each point in the box (they are identical in the center of the box). Solutions indicate which types of firms are active in equilibrium (termed the "regime") and the pattern of foreign affiliate sales and exports. They show that affiliate production is most important when the two countries are similar in size and in relative endowments. The intuition is found by considering what happens when the two countries are quite dissimilar in either of these dimensions. If one country is quite large relative to the other, the dominant firm type will be single-plant firms located in that country serving the small country with exports: it does not pay to incur a plant-specific fixed cost in a small market. If the two countries are quite dissimilar in relative factor endowments, the dominant firm type will be single-plant national firms located in the country which is abundant in the factor used intensively in the multinational's industry.

An integrated treatment of horizontal and vertical multinationals was developed in Markusen's knowledge-capital model (Markusen 2002). Specifically, the model rests on three assumptions relating to knowledge-based assets. First, the services of these assets are easily used in foreign production facilities (transportability or fragmentation). Second, the production of knowledge-based assets is skilled-labor intensive (skilled-labor intensity). Third, knowledge-based assets can yield their full productivity in multiple locations at the same time (jointness). The first two properties support vertical firms while the third supports horizontal multiplant firms.

The contribution of this model is that it yields clear, testable predictions about how foreign affiliate production and trade should be related to the size, size differences, skilled and unskilled labor endowments, and trade and investment cost barriers for two countries. An important extension to include physical capital as a third factor and a third country is found in Bergstrand and Egger (2007).

Empirical Evidence Once again, a good deal of empirical evidence is found in Caves (2007) and Markusen (2002). Brainard (1997) gave convincing evidence for the first time that FDI is not closely related to factor endowments (further discrediting the simple cost-of-capital approach), but much more closely related to country similarity. The ratio of foreign affiliate production to home exports to the host market is increasing in trade costs, increasing in corporate scale economies, and decreasing in plant

scale economies. Only that proportion of affiliate production that is destined for export is related to factor endowments, but that makes sense given the vertical motive for foreign production.

Carr, Markusen, and Maskus (2001) found strong support for the knowledge-capital model. In subsequent work, Markusen and Maskus (see Markusen 2002) found strong support for the horizontal model and virtually no support for a pure vertical model, and could not reject the pure horizontal model in favor of the more complex knowledge-capital model.

Combined with the Brainard paper, these results give strong confirmation to simple summary statistics that the vertical model, the most natural incorporation of the cost-of-capital approach, is a very poor fit indeed. Somewhat later and more sophisticated work by Braconier, Norbäck, and Urban (2006) has discovered more evidence in favor of vertical production, and has found it where the knowledge-capital model suggests it should be found. The addition of physical capital and a third country in Bergstrand and Egger (2007) clears up a number of issues. Finally, readers can find an empirical analysis of the horizontal model extended to heterogeneous firms in Helpman, Melitz, and Yeaple (2004).

The Way Forward There seems no clear direction to the research agenda at this time. Both theoretical and empirical work on multicountry models is needed. Most affiliate activity is directed at local markets as we have noted, and for this part of FDI studying bilateral relationships is fine. But the phenomenon of export-platform production is quantitatively important, and there are only a few papers on this topic. Competition among host countries for inward investment is also interesting and important for development economics.

Research integrating models of internalization/outsourcing into the ownership/location models would be valuable. There is some clear link in models that focus on knowledge capital: the same property that makes knowledge easy to transfer makes it easily dissipated through agent opportunism. There is much less of a link with some newer literature on the property-rights, hold-up approach.

While much has been done on country characteristics (size, endowments, trade and investment costs) as determinants of FDI, somewhat less has been done on the converse question: the effects of inward FDI on host countries. There certainly has been good work on the effect of inward FDI on local labor markets and local firms, but this literature remains disjoint from much of what is covered here. The relationship between host-country governments and multinationals deserves more work. Weak results on taxes as determinants of FDI may reflect the fact that, while multinationals don't like taxes, they value strong physical, educational, and institutional infrastructure that taxes bring.

A few papers have interfaced the so-called New Economic Geography with the theory of the multinational. It seems clear that results from the geography literature on the instability of diversified (or dispersed) equilibria in the presence of moderate trade costs break down when multinational firms are added: horizontal multinationals arise precisely when countries are similar and trade costs are moderate to high, and equilibria in the presence of horizontal multinationals are stable. But much more remains to be done. Similarly, the strategic-trade policy literature focuses almost exclusively on single-plant nationally-owned firms, yet it is precisely in those industries in which scale economies and imperfect competition dominate that we find multinationals.

Individuals working in theory would do well to revisit the international business literature as indicated above. That literature, while frustratingly informal and lacking in testable hypotheses, contains a great richness of insights and data that should be fertile ground for new theoretical ideas.

See also foreign direct investment: the OLI framework; internalization theory; knowledge-capital model of the multinational enterprise; multinational enterprises; outsourcing/offshoring

FURTHER READING

Bergstrand, Jeffrey H., and Peter Egger. 2007. "A Knowledge and Physical Capital Model of International Trade Flows, Foreign Direct Investment, and Multinational

Enterprises." *Journal of International Economics* 73 (2): 278 308. Resolves some empirical puzzles by adding a third factor (physical capital) and a third country to the knowledge capital model.

Braconier, Henrik, Pehr Johan Norbäck, and Dieter Urban. 2005. "Reconciling the Evidence on the Knowledge Capital Model." *Review of International Economics* 13 (4): 770 86. Finds evidence of vertical FDI, previously lacking, consistent with the knowledge capital model by using new data and a new estimation strategy.

Brainard, S. Lael. 1993. "A Simple Theory of Multinational Corporations and Trade with a Trade off between Proximity and Concentration." NBER Working Paper No. 4269. Cambridge, MA: National Bureau of Economic Research. Theory paper analyzing whether firms choose exporting or two plant (horizontal) strategies in a two country model, similar to Horstmann and Markusen (1992).

———. 1997. "An Empirical Assessment of the Proximity Concentration Trade off between Multinational Sales and Trade." *American Economic Review* 87 (4): 520 44. Important paper formally documenting for the first time that foreign affiliate sales are increasing in country similarity and not in factor endowment differences: a death knell for the cost of capital approach.

Carr, David L., James R. Markusen, and Keith E. Maskus. 2001. "Estimating the Knowledge Capital Model of the Multinational Enterprise." *American Economic Review* 91 (3): 693 708. Estimates and finds good support for Markusen's knowledge capital model.

Caves, Richard E. 2007. *Multinational Enterprise and Economic Analysis*. 3d ed. Cambridge: Cambridge University Press. A rich set of theoretical ideas and empirical evidence; excellent background reading for both theoretical and empirical researchers.

Dunning, John H. 1973. "The Determinants of International Production." *Oxford Economic Papers* 25 (3): 289 336. Classic early paper in the international business tradition, contains the roots of the OLI approach.

Helpman, Elhanan. 1984. "A Simple Theory of Trade with Multinational Corporations." *Journal of Political Economy* 92 (3): 451 71. Early formal model of vertical FDI, single plant firms that geographically separate head quarters and plant.

Helpman, Elhanan, Marc Melitz, and Stephen Yeaple. 2004. "Exports versus FDI with Heterogeneous Firms." *American Economic Review* 94 (1): 300 16. Heterogeneous firm approach to horizontal affiliate production: more productive firms choose FDI over exporting.

Horstmann, Ignatius, and James R. Markusen. 1992. "Endogenous Market Structures in International Trade." *Journal of International Economics* 32 (1 2): 109 29. Theory paper analyzing whether duopolists choose exporting or two plant (horizontal) strategies in a two country model, similar to Brainard (1993).

Markusen, James R. 1984. "Multinationals, Multi Plant Economies, and the Gains from Trade." *Journal of International Economics* 16 (3 4): 205 26. Early formal model of horizontal FDI in a model with identical countries: FDI supported by firm level economies of scale due to joint input nature of knowledge based assets.

———. 2002. *Multinational Firms and the Theory of International Trade*. Cambridge, MA: MIT Press. Formal theory and empirical estimation of MNE models; concentrates somewhat on defining, developing, and estimating the knowledge capital model.

Markusen, James R., and Anthony J. Venables. 1998. "Multinational Firms and the New Trade Theory." *Journal of International Economics* 46 (2): 183 203. General equilibrium theory of horizontal production in an oligopoly framework where countries can differ in size and/or relative endowments.

———. 2000. "The Theory of Endowment, Intra Industry, and Multinational Trade." *Journal of International Economics* 52 (2): 209 34. General equilibrium theory of horizontal production in a monopolistic competition framework; stresses the role of trade costs and relates results to the New Economic Geography.

JAMES R. MARKUSEN

■ foreign direct investment and exit of local firms

The exit of local firms refers to the discontinuing of operations by firms owned by local shareholders (i.e., not affiliates of foreign-owned multinationals, which represent foreign direct investment, or FDI). This

cessation of activities might involve firms closing down completely and exiting the industry, or firms discontinuing their current operations but continuing activities in other guises (e.g., under a new name in the same or a different industry). The former situation is the one more frequently studied and is the focus of this entry. Empirical data show that firm exit (in the sense of complete shutdowns) is a fairly frequent phenomenon. In a comprehensive overview paper Caves (1998) provides evidence that in the United States around 7 percent of all firms shut down annually, accounting for about 3 percent of market share in the economy.

Economists have long debated whether there is a link between the large and growing activities of foreign multinationals and the exit of local firms in host countries (e.g., Lall 1978; McAleese and Counahan 1979). The question became even more topical at the end of the 20th century with the increasing concern about globalization and the fears this has sparked for domestic economic activity, in particular the jobs associated with it (Barba Navaretti and Venables 2004). There are three interrelated yet distinct parts to this question: first, whether there are differences in exit patterns between affiliates of foreign-owned multinationals and local firms; second, whether a foreign acquisition affects exit probabilities of the local takeover target; and third, whether the presence of foreign-owned multinationals affects, either positively or negatively, the exit and survival prospects of unrelated local firms.

Comparison of Exit in Foreign and Local Firms
When comparing the incidence of shutdown across foreign affiliates and local firms one may subscribe to one of three views. The first contends that affiliates of multinationals have shallower roots in the domestic economy than local firms and therefore are more likely to shut down production and move to another location, especially when the economy is hit by a negative shock (such as, for example, an economic downturn). A second perspective is that multinationals incur substantial costs for investing abroad, not all of which can be retrieved if they leave; hence, affiliates of multinationals are more likely than domestic firms to stay in the economy in order not to

lose the sunk investment costs. A third view contends that both affiliates of multinationals and local firms are purely profit-maximizing operations and, therefore, there is no strong case that either should behave differently from the other.

Analyzing these possible differences in the incidence or probability of exiting is hampered by a number of difficulties. Perhaps the most important of these is the issue of choosing the right comparison group. Affiliates of multinationals are, on average, quite different from the typical local firm — they are larger, more productive, more skill-intensive, and more technology-intensive. Firms that have these attributes are generally less likely to exit than others, irrespective of the nationality of their ownership. Then the question becomes whether one should compare the incidence of shutdown among foreign affiliates with that of an average local firm, or with that of a local firm that is as similar as possible to the average foreign affiliate.

A study by Görg and Strobl (2003a), using data for the Republic of Ireland, illustrates this issue. In a comparison of the total incidence of exit for all foreign affiliates and all local firms they find that local firms are, on average, significantly more likely to cease operations. This finding would, hence, support the optimistic view described earlier. It does not, however, take account of the fact that affiliates of foreign multinationals based in Ireland are, for example, larger and in more technology-intensive industries than the "typical" local firm and the difference in exit rates could be due to only these differences, rather than to the nationality of ownership per se. To deal with this issue, Görg and Strobl use an econometric technique that allows them to calculate the effect of nationality on the exit rate, abstracting from the effect of firm size and characteristics of the industry. When doing that, they find that affiliates of foreign multinationals are in fact more likely to exit than local firms. A similar analysis is undertaken by Bernard and Sjöholm (2003) for affiliates of foreign-owned multinationals in Indonesia, with results that are strikingly similar in qualitative terms, if not magnitude. Comparing foreign multinationals to a typical local firm shows that the former are less

likely to exit. This difference can be explained by firm characteristics other than nationality, however. Once these are controlled for using econometric techniques, affiliates of foreign multinationals are found to be more likely to exit than a comparable local firm.

Foreign Acquisition and Exit These differences between affiliates of foreign multinationals and local firms raise the question of what would happen if a domestic firm were taken over by a foreign multinational. Would one expect these takeover targets to become more or less likely to face shutdown? Theoretically, the result could be either way. First, since multinationals are generally found to be more likely to exit, a change in ownership may also lead to a higher incidence of plant shutdown for the domestic takeover targets. Second, foreign acquisition may be a device to acquire market access, distribution channels, skills, and so on for a new foreign market entrant. Once these resources/capabilities have been ingested, the acquiring firm may divest itself of the acquired establishment and source its requirements from its plants elsewhere. Foreign acquisition may also lead to a lower threat of exit, however, if the foreign acquirer transfers technology, knowledge, or skills to the acquired plant and hence contributes to an improvement in its performance.

Empirical research into this issue poses some challenges, most important that of constructing a valid counterfactual. While one observes what happens to a local takeover target once it has been acquired by a foreign owner, the important question is what would have happened otherwise in order to identify the effect the acquisition has had. Of course, the latter scenario is unobservable as it did not materialize. Advances in statistical and econometric techniques allow researchers to deal with this problem in various ways, and most recent empirical work uses econometric techniques allowing one to essentially compare local takeover targets with a carefully chosen group of similar local firms that have not been taken over. Based on such econometric techniques, findings for the United States and the United Kingdom using data for the 1980s and 1990s suggest that a takeover by a multinational

increases the chances that the takeover target will be closed (see Bernard and Jensen 2007; Girma and Görg 2004).

Impact of FDI on Exit of Local Firms The presence of foreign firms in an economy can also have an impact on the exit performance of unrelated local firms. The early pessimistic view is that foreign multinationals increase competitive pressure in the economy and therefore make it more difficult for local firms to survive, hence increasing the threat of shutdown for these firms. For a given market size, expansion of output by foreign affiliates attracts demand away from local firms. This shift will cause these domestic firms to cut their production, which, if they face fixed costs, will raise their average cost, reduce profitability, and therefore ultimately make them more vulnerable to shutdown. In fact, regardless of cost structure, increased production by foreign affiliates may lead to a reduction in output price (at least in the short run), which will reduce profitability for local firms and, thus, increase their exit threat.

Although this argument is still powerful, a more recent alternative view takes a more benevolent stance on foreign multinationals and asserts that foreign affiliates may transfer some of their technology to unrelated local firms (either voluntarily or involuntarily). This technology transfer allows local firms to improve their production processes and produce at lower average cost. Hence technology transfer can improve domestic firms' performance by lowering cost and increasing profitability and, thus, also lower the chance of shutdown. This positive effect provides a counterbalance to the possible negative influence through increased competition, and the net impact of foreign affiliates on the exit of local firms may hence be positive, negative, or neutral (the last if the positive and negative effects offset each other).

Empirical evaluation of this issue also is hampered by the challenge of constructing a valid counterfactual. One is able to observe what happens to a local firm in a given industry with a certain level of foreign affiliates operating in it; however, what is unobservable is what would have happened in the absence of

these. Most recent empirical work uses econometric techniques to compare local firms in industries with different levels of presence of foreign affiliates, controlling for possible differences in other firm and industry characteristics in order to deal with this problem.

A study of Ireland by Görg and Strobl (2003b) finds that the presence of foreign affiliates has an exit-reducing effect only on local plants in high-tech industries, while it has no impact on local firms in low-tech industries. This suggests that in the former industries, domestic firms benefit from foreign firms through technology transfers that allow them to improve their performance and reduce the threat of exit. By contrast, if anything the results for firms in low-tech industries indicate that the negative and positive effects offset each other. Whether these findings hold for other countries is an issue for further research.

Policy Implications Foreign direct investment is an important feature of the globalized world economy. The impact of FDI on economic development, in particular the interplay between local firms and foreign affiliates of the countries involved, is an important issue for policymakers. Although studies have identified general trends and possible effects, it remains to be seen what lessons can be learned from these for particular economies at different levels of development.

See also agglomeration and foreign direct investment; exchange rates and foreign direct investment; footloose production; foreign direct investment and labor markets; foreign direct investment under monopolistic competition; foreign direct investment under oligopoly; outsourcing/offshoring; subsidies and financial incentives to foreign direct investment; transfer pricing

FURTHER READING

Barba Navaretti, Giorgio, and Anthony J. Venables. 2004. *Multinational Firms in the World Economy.* Princeton, NJ: Princeton University Press. Chapters 7 and 8 provide a textbook treatment of the effect of FDI on local firms, also looking at exit of local firms.

Bernard, Andrew, and Bradford Jensen. 2007. "Firm Structure, Multinationals, and Manufacturing Plant Death." *Review of Economics and Statistics* 89: 193 204. Looks empirically at firm exit in the United States and the role of multinationals and ownership change.

Bernard, Andrew, and Fredrik Sjöholm. 2003. "Foreign Owners and Plant Survival." NBER Working Paper 10039. Cambridge, MA: National Bureau of Economic Research. An empirical examination of differences in exit rates between foreign affiliates and local firms for Indonesia.

Caves, Richard E. 1998. "Industrial Organization and New Findings on the Turnover and Mobility of Firms." *Journal of Economic Literature* 36: 1947 82. A very comprehensive review of the research on firm entry and exit.

Girma, Sourafel, and Holger Görg. 2004. "Blessing or Curse? Domestic Plants' Survival and Employment Prospects after Foreign Acquisition." *Applied Economics Quarterly / Konjunkturpolitik* 50 (1): 89 110. Empirical study of effects of foreign acquisitions on exit of local takeover targets using UK data.

Görg, Holger, and Eric Strobl. 2003a. "Footloose Multinationals?" *The Manchester School* 71: 1 19. An empirical examination of differences in exit rates between foreign affiliates and local firms for Ireland.

———. 2003b. "Multinational Companies, Technology Spillovers, and Plant Survival." *Scandinavian Journal of Economics* 105 (4): 581 95. An empirical investigation of the impact of foreign multinationals on exit of local firms using data for Ireland.

Lall, Sanjay. 1978. "Transnationals, Domestic Enterprises, and Industrial Structure in Host LDC's: A Survey." *Oxford Economic Papers* 30 (2): 217 48. An early and very comprehensive survey of host country effects of foreign multinationals.

McAleese, Dermot, and Michael Counahan. 1979. "Stickers or Snatchers? Employment in Multinational Corporations during the Recession." *Oxford Bulletin of Economics and Statistics* 41: 345 58. An early discussion of the "footloose" nature of affiliates of foreign multinationals in host countries.

HOLGER GÖRG

■ **foreign direct investment and export performance**

The foreign direct investment export link focuses on the role of foreign direct investment (FDI) in export performance. FDI represents international capital flows through which a firm in one country (the home) creates a subsidiary or acquires control over a business enterprise in another country (the host). The firm with FDI is thus a multinational enterprise (MNE), which must have firm-specific advantages (technology or brand name) and choose a host country with location advantages (cheap labor or growing markets). Export performance has many facets, including export volume, technological composition, growth rate, and global competitiveness. FDI plays a significant role in world exports. Broad estimates indicate that MNEs account for two-thirds to three-quarters of world exports, and foreign affiliates of MNEs account for more than one fifth of world exports and one third of developing country exports (UNCTAD 2002). The paucity of data, however, makes quantitative assessments of the effect of FDI on export performance difficult.

Theoretical studies generally identify two types of FDI: vertical and horizontal (Markusen 1995). A manufacturing MNE usually engages in three activities: firm activities producing headquarters services; upstream-plant activities producing components; and downstream-plant activities of assembling. Headquarters with firm activities are located in home countries, while cross-country relocation of downstream- and/or upstream-plant activities results in vertical and horizontal FDI.

Vertical (efficiency-seeking) FDI is associated with relocation of labor-intensive downstream-plant activities to host countries with cheap labor, while in horizontal (market-seeking) FDI, another set of production facilities (including both upstream- and downstream-plant activities) is built in host countries to serve local markets. Motivated by cheap labor and other resources, the vertical FDI firm may export capital-intensive components and headquarters services to its host-country affiliates for assembling, and ship finished products back to the home country or third-country markets (Helpman and Krugman 1985). Horizontal FDI tends to overcome trade barriers such as tariffs and transportation costs in order to maintain or expand market shares in host countries (Markusen 1984).

The sharpness of the distinction between the two types of FDI may be best explained by MNEs in semiconductors and automobiles. In the 1970s, U.S. chip firms made substantial investments in assembly and testing facilities in Southeast Asia, motivated by cheap labor there. Semiconductors were shipped from the U.S. design and manufacturing complexes to the assembly and testing facilities in Southeast Asia, and then shipped back to sales destinations in the United States and elsewhere. In contrast, Japanese automakers in the 1980s established production lines in the United States in order to avoid local protection threats by replacing existing exports with local production.

How Does FDI Affect Home-Country Exports?
FDI affects home-country export performance through direct effects on trade as well as indirect effects through various channels. The direct effects depend to a large extent on the type of FDI. Vertical FDI could enhance the home country's exports of intermediate products (parts and components) required for assembling. This intrafirm trade is called a complementary effect of exports. Horizontal FDI may lead to an increase in exports of capital goods and intermediate products in the short run, but its long-term impact turns out to be export reduction (so-called substitution effects).

Empirical evidence from developed countries (e.g., the United States, Japan, and Sweden) indicates the FDI-export link to be one of complements at country and industry levels, but of "substitution" at the product level (Blonigen 2001). When activities of foreign affiliates can be classified as vertical and horizontal, evidence is consistent with the theoretical prediction that vertical FDI complements exports and horizontal FDI substitutes for exports. FDI from high-income developing economies (e.g., Hong Kong, Korea, Singapore, and Taiwan) to other developing countries (e.g., China) has been found to be a contributory factor

for expanding home-country exports (UNCTAD 2002).

In the long run, FDI may indirectly affect home-country exports by improving the competitiveness of parent companies; upgrading industrial structure; and creating spillovers to the rest of the home economy.

How Does FDI Affect Host-Country Exports? Effects of FDI on the host country's exports may differ, depending on the MNE's motives. Horizontal FDI is oriented primarily to the host-country market, so it does not directly contribute to host-country exports. Vertical FDI directly boosts host-country exports, since output resulting from FDI is typically intended for export. Most vertical FDI is hosted in developing countries in labor-intensive industries such as textiles, garments, and home electronics. China, for example, attracted a significant amount of vertical FDI, which contributed 57 percent of its total exports in 2004 (Zhang 2006). While the overall impact is export-promoting, the real increase in host-country exports depends on the local content of value-added with exports. If local firms supply intermediate products to MNE affiliates, host-country exports would be much greater than in cases where the inputs are imported from outside the host economy.

A popular form of vertical FDI is located in export processing zones (EPZs), in which materials are imported duty free and transformed for exports, with strictly controlled trade with the rest of the host country. EPZs allow exploiting the location-specific assets of a host country while avoiding the restrictions imposed by its trade regime, providing good infrastructure, and offering fiscal incentives. Many countries (such as Costa Rica, China, Haiti, Mauritius, Bangladesh, Singapore, Malaysia, and Sri Lanka) have enjoyed spectacular growth in manufactured exports from EPZs. The most successful exports have been garments and semiconductors.

Theories and evidence suggest that FDI may have several potential mechanisms to influence host export performance. Its contribution may derive from additional capital, technology, and managerial know-how; training for the local workforce; and access to global and especially home-country markets. These resources and market access may complement a host country's own resources and capabilities, and may provide some of the missing elements for greater export competitiveness. The host country may build on these to enter new export activities and improve its performance in existing ones (UNCTAD 2002). Many studies emphasize the role of FDI in building host export competitiveness in manufacturing goods. Three positive effects of FDI may be identified: helping transfer of technology and new products for exports; expanding market access for exports; and building dynamic comparative advantages. On the other hand, some negative effects of FDI on host exports have been observed. FDI may (1) lower or replace domestic savings and investment; (2) transfer technologies that are low level or too complicated for the host country to absorb due to lack of human capital; (3) target primarily the host country's domestic market and thus not increase exports; (4) inhibit the expansion of indigenous firms that might become exporters; and (5) neglect the host country's dynamic comparative advantages by focusing solely on local cheap labor and raw materials (Zhang 2006).

While potential benefits of FDI to host exports exist, they do not automatically accrue. To what extent a host country can capture them depends largely on its own strategies and efforts. Opening up to FDI is only the first step, and host countries need to realize existing comparative advantages based on initial capabilities. For countries with weak industries and exports, FDI may well lead only to a short-lived boost in export performance. To build a more sustainable and dynamic export base, host countries have to use proactive policies, including selective liberalization to reconcile efforts to attract vertical FDI with the need to protect particular industries; attracting FDI by targeting investment conducive to export competitiveness and upgrading; and strengthening domestic enterprises, as well as the skills, capabilities, and institutions necessary for successful export-oriented strategy.

See also appropriate technology and foreign direct investment; export processing zones; intrafirm trade; knowledge-capital model of the multinational enterprise;

linkages, backward and forward; technology spillovers; vertical versus horizontal foreign direct investment

FURTHER READING

Blonigen, Bruce A. 2001. "In Search of Substitution be tween Foreign Production and Exports." *Journal of In ternational Economics* 53 (1): 81 104. An empirical study of substitution and complement effects of FDI on home country exports, using product level data.

Helpman, Elhanan, and Paul Krugman. 1985. *Market Structure and Foreign Trade*. Cambridge, MA: MIT Press. Chapters 12 and 13 offer a central theoretical analysis of vertical FDI and its link to trade.

Markusen, James. 1984. "Multinational, Multiplant Econ omies, and the Gain from Trade." *Journal of Interna tional Economics* 16(3 4): 205 26. An important work on horizontal FDI.

———. 1995. "The Boundaries of Multinational En terprises and the Theory of International Trade." *The Journal of Economic Perspectives* 9 (2): 169 89. A thor ough review of the research on FDI and international trade.

United Nations Conference on Trade and Development (UNCTAD). 2002. *World Investment Report 2002: Transnational Corporations and Export Competitiveness*. New York: United Nations. Each year UNCTAD spotlights an important FDI issue; the 2002 report focused on empirical evidence about effects of FDI on exports in host countries, especially developing economies.

Zhang, Kevin Honglin. 2006. "The Role of FDI in China's Export Performance." In *China as the World Factory*, edited by Kevin H. Zhang. London: Routledge, 139 55. A comprehensive analysis of positive and negative effects of FDI on exports from China, a nation hosting the largest amount of FDI in the developing world and being globally the second largest exporter in 2007.

KEVIN HONGLIN ZHANG

■ foreign direct investment and innovation, imitation

Following successful innovation of a new process or product technology, production can be shifted abroad through a variety of channels: firms can li cense their technologies to foreign firms, undertake foreign direct investment (FDI), or may find that their technologies are transferred abroad through imitation by foreign firms. With FDI, the incumbent firm establishes a plant in another country, whereas with imitation, production in the other country is done by a rival firm (or firms).

FDI, innovation, and imitation each typically requires the investment of resources. Most innova tion results from intentional efforts to develop new varieties of products, higher-quality levels of existing products, or lower-cost methods of production (process improvements). These efforts require firms to bear significant costs with an uncertain outcome. While there may be some products for which imi tation is straightforward, in most cases imitation, like innovation, is also a costly and uncertain activity. Mansfield, Schwartz, and Wagner (1981) find that imitation costs average two-thirds of innovation costs for firms in the chemical, drug, electronics, and machinery industries. Undertaking FDI also incurs costs (prior to production), such as fixed costs of establishing new plants, adapting technologies to new economic environments, finding local suppliers, and the like.

Each of these activities FDI, innovation, and imitation must yield a reward in the form of an expected stream of profits sufficient to offset the initial costs or else firms would have little incentive to undertake them. Properly discounted, the total ex pected profits from an innovation must at least cover the costs of generating the innovation for firms to be willing to invest resources in the innovation. As the costs of a successful imitation tend to be lower than for the original innovation, the profits of a successful imitator can be lower (or shorter in duration) than for an innovator, but again imitators need to rewarded, through the expected profits from a successful imi tation, for their imitation expenses. For FDI, the expected profits must rise by enough to compensate for the up-front costs of shifting technology abroad.

Additionally, these activities may be constrained by the availability of factors such as skilled labor. In countries where firms innovate, there must be enough resources for both R&D and production

uses; in countries where firms imitate, there must be enough for imitation efforts, production by successful imitators, and production by multinational firms. Thus FDI, innovation, and imitation are all affected by incentives and resource availability the two key factors that determine practically all economic activity.

How FDI Affects Innovation The costs of innovation need to be offset by the expected stream of profits from selling the product. As FDI is undertaken to enhance profits, it would seem that the opportunity to undertake FDI should accelerate innovation by enhancing the reward for innovation.

Suppose FDI involves shifting production to a lower-cost location. If the opportunity to shift production abroad arrives at a higher probability, then the higher profits under FDI should begin sooner, all else being equal. Glass and Saggi (1999), however, show that the effect of faster arrival of FDI opportunities on innovation can be essentially zero due to increased exposure to imitation. The faster arrival of higher profits under FDI is offset by a shorter duration of profits before imitation, when FDI opportunities are beyond the control of firms.

Why then are FDI opportunities commonly believed to stimulate innovation? Perhaps FDI occurs not in response to increases in opportunities to shift production abroad, but rather in response to other influences. Instead of a faster arrival of fixed FDI opportunities, suppose that FDI increases in response to variations in the underlying economic environment. For example, suppose the labor supply abroad increases to provide more labor for production there. In response, FDI increases to shift production from home to abroad, and innovation also increases, as shown in Glass and Saggi (2002). Other changes, such as a reduction in the difficulty of adapting technologies for production abroad, can also increase FDI and innovation. Crisuolo, Haskel, and Slaughter (2005) provide evidence that firms that are globally engaged do innovate more.

How FDI Affects Imitation By shifting production abroad, FDI typically generates technology spillovers through demonstration effects. More can be learned about a technology when it is produced

locally than when only the final product can be seen. Although inspecting the final product can yield useful insights, full knowledge of the production process is hard to determine from the end result. When production occurs locally, however, workers at the multinational firm observe the production process firsthand. Workers who have been exposed to the technology, and perhaps specially trained, may leave the multinational to work for local firms or even start their own firms. Local suppliers of intermediates may also serve as sources of information. These knowledge flows act to lower the cost and difficulty of imitation. Hence, by reducing the cost of imitation, FDI should encourage imitation through incentive effects.

On the other hand, FDI vies with local production for local resources such as skilled labor. In some circumstances, FDI might deter imitation if multinationals bid up prices for scarce resources. Furthermore, when FDI is motivated by cost savings, multinationals have lower production costs than firms that export the good from elsewhere, so the profit margins earned when producing an imitation of a multinational's product should be less than that earned when producing an imitation of a product produced elsewhere. Due to the reduced profit margins for imitators, FDI could deter imitation. Glass and Saggi (1999) show that a faster arrival of opportunities to shift production abroad through cost-saving FDI can leave the aggregate rate of imitation essentially unchanged. So even though FDI makes imitation easier, there is no guarantee that FDI spurs local imitation.

How Innovation Affects FDI Multinational firms need to possess an ownership advantage in order to overcome the inherent difficulties of operating in multiple countries. Innovation yields the technological expertise that often provides the source of ownership advantage for multinational firms. As innovation is, in a sense, a necessary precondition for FDI, innovation should spur FDI.

Who is conducting the innovation matters, however. Doing more innovation should lead the innovating firm to do more FDI. But to the extent that innovation is being done by rival firms, FDI may be

deterred. Innovation by rivals poses a risk that the profit stream that rewards innovation may be terminated (or reduced) by a successful innovation by another firm. In some industries, the existing leaders have significant advantages from successful past innovations and therefore have a greater incentive to innovate, whereas in other industries, firms are on more equal footing and innovation is spread across many firms. This distinction suggests that the relationship between innovation and FDI may differ across industries.

For the newest technologies, firms usually opt for FDI over licensing on account of large transaction costs that arise due to the presence of asymmetric information. At the moment when a new technology is first created, only the innovating firm is well informed about its key attributes. Thus transferring such technologies to independent firms in other countries via arm's-length contracts might be rather difficult since such firms would typically not have a reliable estimate of the value of the technology. Under such a scenario, the innovating firm might choose to undertake FDI by establishing a fully owned subsidiary as opposed to licensing the new technology to a foreign firm. By impeding arm's-length contracting, transaction costs can induce internalization on the part of an innovating firm.

Furthermore, such internalization might also be motivated by strategic considerations: a firm may be unwilling to share its newest technologies with foreign firms that could become future competitors. It may be difficult, or perhaps even impossible, to prevent a licensee from terminating the contract and undertaking independent production after it has mastered the new technology. A further worry is that a current licensee might use the technology to invent a future technology that would make it an even fiercer competitor. Both the transaction cost perspective and strategic considerations suggest that firms will typically transfer their newest technologies through FDI and relatively more mature, and less valuable, technologies through licensing and other arm's length arrangements. A wealth of econometric and case-study evidence supports this hypothesis.

How Imitation Affects FDI When FDI brings about higher profits through costs savings, imitation poses the risk that the enhanced profit stream under FDI will be terminated. Thus imitation would seem to deter FDI. To determine how imitation affects FDI incentives, however, Glass and Saggi (2002) have argued that, as multinationals need not be the only firms targeted by imitation, the relative risk of imitation for a multinational must be compared to the imitation risk of producing in the home country. Even though FDI lowers the cost of imitation, it also lowers the multinational's production costs and therefore the profit margin earned by an imitator targeting the multinational's product. So there is reason for both types of imitation to exist: easier imitation targeting multinationals and more difficult imitation (that is more richly rewarded in terms of profits) targeting products produced elsewhere. When imitators target technologies regardless of whether FDI has occurred, increased risk of imitation need not deter FDI if the imitation risk also increases for firms that produce back in their home countries. What matters is how much more exposed to imitation multinationals are relative to firms that export. Glass and Saggi (2002) demonstrate a tendency for the relative risk of multinationals to remain constant, so the increased risk of imitation for a multinational does not vary as many other parameters of the model change.

For example, if imitation becomes more difficult, the rate of imitation falls, which one might think would encourage FDI. But both imitation targeting multinationals and imitation targeting nonmultinationals decrease, and they decrease in equal proportion, so the relative imitation risk for multinationals is essentially unchanged. In addition, there are effects operating though the labor constraint: due to the increased difficulty of imitation, the labor needed for imitation can increase, even though imitation falls. The increased use of labor for imitation crowds out FDI, as there is less labor left for production by multinational firms. In this case, less imitation (due to increased difficulty) leads to less FDI.

Imitation can be thought to encourage FDI in some settings. Glass and Saggi (1998) construct a

model with two types of FDI: FDI that transfers state-of-the-art technology and FDI using older technology. In that model, successful imitation of the older technology is required before FDI transferring state-of-the-art technology becomes feasible. The initial imitation is needed to establish a knowledge base in order for the technology transfer costs for the new technology to become reasonable. In this setting, although imitation of the latest technology could deter FDI, some initial imitation of older technologies is necessary for more advanced FDI to occur.

The impact of imitation on FDI can also depend on whether imitation is varied or whether imitation changes in response to other parameters of the model. Glass and Wu (2007) contrast the results for the effects of imitation on FDI when imitation changes on its own or in response to changes in other parameters of the model, as well as depending on whether imitations are quality improvements or new varieties, and whether innovations are done by leaders or rival firms (followers). Mostly, imitation and FDI move together in quality ladder models, except when innovations are done by followers. Increases in imitation may decrease FDI when innovations are new varieties (see Helpman 1993; Lai 1998). There are also models with R&D having decreasing returns to scale (see Dinopoulos and Segerstrom 2005; Sener 2006).

In sum, the interrelationships between FDI, innovation, and imitation are quite complex. Claims that may seem intuitively obvious, such as that innovation spurs FDI or even that FDI spurs innovation, need not be generally true. Similarly, FDI need not spur imitation and imitation need not always deter FDI. Given the many possible scenarios uncovered by theory, there is a great need for empirical work to sort out which outcomes are predominant. A better sense of what is happening in the data will lead to a better understanding of how imitation, FDI, and innovation matter to firms.

See also agglomeration and foreign direct investment; foreign direct investment under oligopoly; intangible assets; intellectual property rights and foreign direct investment; multinational enterprises; technology licensing; technology spillovers; trade-related investment measures (TRIMs)

FURTHER READING

Crisuolo, Chiara, Jonathan E. Haskel, and Matthew J. Slaughter. 2005. "Global Engagement and the Innovation Activities of Firms." NBER Working Paper No. 11479. Cambridge, MA: National Bureau of Economic Research. Evidence that globally engaged firms innovate more.

Dinopoulos, Elias, and Paul Segerstrom. 2005. "Multinational Firms and Economic Growth." Mimeo. University of Florida and Stockholm School of Economics. Compares effects of policies that encourage multinational firms to effects of globalization in the form of geographic expansion, in a scale invariant growth model.

Glass, Amy J., and Kamal Saggi. 1998. "International Technology Transfer and the Technology Gap." *Journal of Development Economics* 55 (2): 369 98. A model of the connection between imitation, FDI, and innovation that distinguishes between the type of technology (new or old) transferred through FDI.

———. 1999. "Foreign Direct Investment and the Nature of R&D." *Canadian Journal of Economics* 32 (1): 92 117. Shows how exogenous increases in opportunities for cost saving FDI affect imitation and innovation.

———. 2002. "Intellectual Property Rights and Foreign Direct Investment." *Journal of International Economics* 56 (2): 387 410. Models the interrelationships between imitation, FDI, and innovation when imitation is determined within the model and innovations are quality improvements.

Glass, Amy J., and Xiaodong Wu. 2007. "Intellectual Property Rights and Quality Improvement." *Journal of Development Economics* 82 (2): 393 415. Models the interrelationships between imitation, FDI, and innovation when imitation is not determined within the model and innovations are quality improvements; contrasts with other findings.

Helpman, Elhanan. 1993. "Innovation, Imitation, and Intellectual Property Rights." *Econometrica* 61 (6): 1247 80. Models the effect of imitation on FDI when innovations are new varieties.

Lai, Edwin L.C. 1998. "International Intellectual Property Rights Protection and the Rate of Product Innovation." *Journal of Development Economics* 55 (1): 133 53. Models the effect of imitation on FDI and innovation when innovations are new varieties.

Mansfield, Edwin, Mark Schwartz, and Samuel Wagner. 1981. "Imitation Costs and Patents: An Empirical Study." *Economic Journal* 91 (364): 907 18. Collects data on imitation cost and time relative to innovations.

Sener, M. Fuat. 2006. "Intellectual Property Rights and Rent Protection in a North South Product Cycle Model." Mimeo. Union College. Includes efforts by successful innovators to deter innovation and imitation by other firms.

Vernon, Raymond. 1966. "International Investment and International Trade in the Product Cycle." *Quarterly Journal of Economics* 80 (2): 190 207. The original product cycle model with FDI.

AMY JOCELYN GLASS AND KAMAL SAGGI

■ foreign direct investment and international technology transfer

A technology gap separates developing countries from the developed world: from manufacturing processes to information technology provision in business, government, and higher education, poorer countries lag behind richer ones. One reflection of this gap is the high concentration of global research and development (R&D) in developed countries: in 2002, developed countries accounted for more than 90 percent of global R&D expenditures (UNCTAD 2005). The transfer of technology from developed to developing countries helps close the international technology gap and contributes to productivity growth in developing countries. Such technology transfer makes much economic sense, since it keeps poorer countries from having to spend already scarce resources on reinventing the wheel.

The results of R&D undertaken in developed countries are transferred globally through many channels, perhaps the most important of which is foreign direct investment (FDI) investment projects whose ownership and control lie in the hands of overseas investors. The establishment of production facilities by foreign companies usually results in the introduction of new production processes and techniques to the local economy, making FDI a major channel of technology transfer. FDI also has come to play an increasingly important role in helping disperse R&D activities globally: during the period 1993 2002, R&D expenditures by foreign subsidiaries of multinational firms increased from $30 billion to $67 billion, representing a rise from 10 percent to 16 percent of global business R&D.

This article examines several questions related to international technology transfer. How do technologies originating in industrialized countries take hold in developing countries? What role does FDI play in this process? What benefits do developing countries enjoy from FDI? And how, if at all, can international organizations such as the World Trade Organization (WTO) facilitate the process of technology transfer?

International Technology Transfer and the Role of FDI While much FDI occurs between industrialized countries, developing countries are becoming increasingly important hosts for such investment flows. In fact, during 2001 5, average net FDI inflows equaled $754.3 billion, of which $212.4 billion (approximately 29 percent) went to developing countries. The global stock of FDI in 2004 stood at approximately $9 trillion with approximately 25 percent of it being in developing countries (UNCTAD 2005).

FDI is undoubtedly one of the primary channels of international technology transfer. It is well known that multinational activity occurs primarily in industries that are characterized by a high ratio of R&D to sales and by large shares of professional, scientific, and technical workers (Markusen 1995). In fact, a basic tenet of the theory of the multinational firm is that such firms rely heavily on their technology-based assets and well-established brand names to offset the logistical disadvantages of operating in multiple counties as well as to successfully compete with local firms that are usually more familiar with the local environment. In 1995, of all transactions in royalty and license fees, transactions within the same firm made up in excess of 80 percent, so most explicit

trade in technology takes place within multinational firms (UNCTAD 1997).

Other private sector channels for technology transfer include technology licensing and joint ventures. With technology licensing, a licensee will pay royalties and fixed fees to a foreign licensor in exchange for the right to use its technology. Joint ventures, on the other hand, are typically organized as partnerships between foreign and local companies. Then there are the so-called indirect routes for technology transfer, which include international trade in goods and services and the movement of labor between countries. Another such route is the imitation or "reverse engineering" of technological products, which essentially involves taking them apart for analysis and then independently reproducing them — a practice that allows local entrepreneurs and companies to replicate products and technologies from abroad without paying any royalties or fees.

The usage of the word *transfer* in the phrase *technology transfer* conveys the impression that the process is smooth and costless. Such is hardly the case. The fact that developing countries lag behind the technology frontier merely creates the potential for technology transfer. For this process actually to occur, providers and acquirers of new technologies have to undertake deliberate and often significantly large investments. Substantial empirical evidence indicates that international technology transfer does not come cheap. Economists regard knowledge (and its application in the form of technology) as a nonrival good, meaning that it can be accessed by more than one person. The nonrival nature of knowledge does not mean, however, that it can be transferred free of cost. Rather, it simply implies that if two people are willing to pay the cost of adopting a new idea or a technology, they can do so without interfering with each other's decisions.

Analyses of FDI must always bear in mind one basic observation: multinational firms engage in the transfer of technology primarily to maximize profits and value for their stockholders. Thus any technology transfer activity that does not contribute to profit-making will typically not be attractive for a multinational company. For example, if transferring a technology runs the risk of strengthening a company's local rivals, the company should be expected to try to thwart the process of technology diffusion to some degree: it may choose not to invest locally and may simply export the product from its home market or may transfer only peripheral technologies as opposed to core ones.

Despite such concerns, FDI often generates benefits for both consumers and local companies in developing countries. Benefits to consumers include access to higher-quality goods or completely new goods that were not produced locally. Benefits to local companies include what is known as the "demonstration effect," in which local firms adopt a technology from overseas after observing its successful introduction by a multinational. People trained or previously employed by multinational firms can provide added benefit to local business by helping with technology transfer when they change jobs or start new firms.

Multinationals also engage in technology transfer to firms that either supply them or buy their products. An example can be found in Mexico's automobile industry. After the North American Free Trade Agreement (NAFTA) was ratified in 1992, U.S. car companies set up manufacturing plants in Mexico. Within five years this had led to hundreds of domestic producers of car parts and accessories springing up (Moran 1998). United States and other multinational corporations transferred technology to these Mexican suppliers, who, as a direct result, gained expertise in industry best practices and quality control. The Mexican story is by no means unique. A rich body of evidence indicates that Mexico's experience has been replicated across the world (Saggi 2002).

Policy Implications The earlier discussion suggests that the process of technology transfer via FDI may not always work smoothly. But does this necessarily create a role for government support or regulation of such activities? Historical experience of countries such as Japan, South Korea, and China suggests that many national governments perceive this to be the case. These countries have considerable experience in attracting and regulating private

sector international technology transfer. Both Japan and South Korea have been deliberately selective in their choice of technology transfer method, preferring technology licensing over FDI. Their experiences have been studied in some depth. It is well known that during the 1960s and 1970s, Japan's Ministry of International Trade and Industry (MITI) limited competition between potential local buyers of foreign technologies to keep the prices they paid for it low. Furthermore, MITI never formally encouraged FDI and instead encouraged local companies to license technology from overseas. It even insisted that foreign firms share their technology with local firms as a precondition for doing business in Japan. South Korea presents a similar story. In fact, discouraging FDI while encouraging technology licensing was a central theme of South Korea's industrial policy.

There are several reasons why such countries may prefer technology licensing to FDI. They may share a fear that FDI may lead to a transfer of what are called "black box" technologies, in which local companies do not learn very much about how the technology was developed. Another perception is that FDI gives the foreign companies that supply the technologies a stronger bargaining position than the local ones receiving it. Governments also believe that FDI has fewer of what are called "spillover" benefits for local businesses, compared with technology licensing and joint ventures between local and foreign firms.

The evidence for these views remains quite thin, however. In fact, several studies have found that the opposite is true. For example, technology licensing may offer fewer opportunities for profit-making than is sometimes perceived. In addition, a licensee may need additional permission from the license holder to export the technology to a new country. Or it may need permission to make additional modifications or improvements. This shows that, if anything, technology licensing can carry a number of restrictions that may hamper technological development and growth of the licensee's businesses.

Because of these limitations, governments in many developing countries have come to adopt a more liberal approach to FDI than in the past. In fact, today most countries compete with one another to attract more FDI by providing a range of fiscal and other financial incentives. This is more relevant for manufacturing than for services, however. For example, in recent years, while China has been the biggest recipient of FDI in the world, its services sector has not been particularly open to FDI. This was, in fact, an important sticking point during China's negotiations regarding its entry into the WTO.

For international technology transfer to work well, policies need to take into account the needs of both buyers and sellers of foreign technology. As we have seen, the point of technology transfer from the perspective of developing countries is to save them from having to reinvent the wheel. At the same time, the costs of new technologies need to be kept low. From a policy point of view, this means that governments need to invest in education, infrastructure, and local expertise to adopt and adapt new technologies.

From the perspective of technology suppliers, transferring technologies to poorer countries comes with risks, which can be mitigated by appropriate (but not excessive) regulation. Historically, developing countries have not always appreciated the fact that suppliers of foreign technology need to profit from their endeavors. But today, the pendulum is in danger of swinging too far in the other direction. Developing countries may have become overeager to attract FDI into manufacturing, even while still restricting it in services. But this is problematic, as policies that favor one sector over another distort the flow of FDI relative to what would occur as a result of purely market forces. Few developing countries have comparative advantage in world-class manufacturing, and favoring the manufacturing sector can deprive the development of the service sector, leading to a misallocation to inward FDI as well as complementary domestic resources.

The Role of International Institutions Most international treaties at the United Nations level make some reference to the need for rich countries to help developing countries meet their treaty obliga-

tions by transferring technologies under favorable conditions. For example, Article 7 of the WTO's Trade-Related Aspects of Intellectual Property Rights (TRIPS) agreement notes that intellectual property protection should "contribute to the promotion of technological innovation and to the transfer and dissemination of technology, to the mutual advantage of producers and users of technological knowledge and in a manner conducive to social and economic welfare, and to a balance of rights and obligations." Article 66 of the same agreement commits developed countries to find ways to encourage international technology transfer to developing countries and to help improve their technological base.

Governments of industrialized countries are less enthusiastic about subsidizing technology transfer, however, in part because they are reluctant to interfere with what is now an activity dominated by the private sector. Their earlier pledges to transfer technology to help developing countries with international agreements on biodiversity and global climate change remained mired in debate at the end of 2007, as did discussion on concrete steps to implement Article 66 of TRIPS.

The WTO has a working group on trade and technology transfer, but through the end of 2007, it had focused primarily on information exchange between representatives of the group's member countries. And even here, the information provided had been little more than what is freely available in the academic literature.

The potential certainly exists for countries in the group with more experience, such as Japan and South Korea, to provide a mentoring role to less developed ones. Developing country members would benefit from practical case studies from Japan, for instance, that show its reasons for adopting certain technology transfer policies, and how these helped to strengthen local industry.

Poorer countries can also benefit from guidance on the structure of international technology transfer contracts, including, for instance, information on reasonable royalty rates, the conditions sellers of technology have been willing to accept, and the types of contract clauses that have proved helpful in encouraging local technological development.

It is a staple of today's development thinking that technology transfer from developed to developing countries is important for generating sustained economic growth. But historical experience shows that, to be innovators themselves, countries must first learn to absorb foreign technologies and adapt them to local conditions.

Because of government intervention, in several Asian countries technologies in the private sector were frequently transferred through technology licensing rather than FDI. This may have allowed some countries such as South Korea to obtain foreign technologies at relatively low prices. Although such policies are not necessarily the most desirable ones, they were more successful than the policies of countries such as India that focused on achieving self-reliance by fostering imitation and adoption of foreign technologies under a regime that protected local industries from foreign competition.

Today's multilateral trading system is a markedly different one, and it seems unlikely that developing countries can adopt the interventionist policies used by even Japan and South Korea, let alone India: several multilateral agreements of the WTO explicitly forbid governments from pursuing such policies. Furthermore, policies supporting technology licensing require an effective bureaucracy, which some of the poorest developing countries lack.

Developing countries are likely to maximize the benefits of technology transfer by investing more in local education and infrastructure development while giving multinational firms relatively unrestricted access to their markets. In particular, policies that favor one sector (such as manufacturing) at the expense of another (such as services) seem counterproductive. As an example, it is worth noting that the development and success of India's software sector has far outpaced its manufacturing sector despite the fact that the Indian government has always been rather keen to promote local manufacturing via the pursuit of a restrictive trade policy. In fact, it seems quite likely that the software sector succeeded because it was able to draw on India's investment in

education (as evidenced by the fact that the Indian Institutes of Technology are considered on a par with the best universities in developed countries) and because the government did not (some would say it could not) restrict international trade in the sector.

In addition, governments in rich countries and international institutions such as the WTO have the potential to do more to encourage technology transfer. Similarly, the governments of advanced developing countries can play a more active role in sharing their own experiences of managing technology transfer from developed countries.

See also Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS); appropriate technology and foreign direct investment; foreign direct investment and exit of local firms; foreign direct investment and innovation, imitation; foreign direct investment and labor markets; intellectual property rights and foreign direct investment; internalization theory; joint ventures; linkages, backward and forward; location theory; multinational enterprises; technology licensing; technology spillovers; trade-related investment measures (TRIMs)

FURTHER READING

Markusen, James R. 1995. "The Boundaries of Multi national Enterprises and the Theory of International Trade." *Journal of Economic Perspectives* 9: 169 89. A rigorous review of the economics of the multinational firm viewed from the perspective of international trade theory.

Moran, Theodore. 1998. *Foreign Direct Investment and Development.* Washington DC: Institute for Interna tional Economics. A rich discussion of the impact of FDI on economic development that draws on several recent national case studies.

Saggi, Kamal. 2002. "Trade, Foreign Direct Investment, and International Technology Transfer: A Survey." *World Bank Research Observer* 17: 191 235. A critical review of the economics literature on international technology transfer.

UN Conference on Trade and Development (UNCTAD). 1992 2005. *World Investment Reports.* New York: United Nations. Annual reports on global flows of FDI published by the United Nations Conference on Trade and Development.

KAMAL SAGGI

■ foreign direct investment and labor markets

Private cross-border capital inflows are made up of foreign direct investment (FDI) and portfolio equity and debt flows. *Direct* investments are defined as those which wield a significant influence on the management of the firm. FDI flows therefore comprise the financing of overseas greenfield investments (where new operational facilities are constructed from the ground up) and further expansions, and of cross-border mergers and acquisitions by multinational enterprises (MNEs). The total direct capital owned by nonresidents in a given country in a particular year constitutes the stock of FDI.

The regional and sectoral destination of FDI flows has changed substantially over the decades. At the outbreak of World War I, more than 80 percent of the foreign capital stock was located in developing economies, reflecting the importance of railway building, the extractive industries, and the colonial control of international trade at that time. In the early 21st century, by contrast, almost 80 percent of the global FDI stock is located in the developed world, and FDI is associated with a different range of activities. The transnationality of an industry today is associated with the sectoral importance of knowledge capital such as "patents, blueprints, formulae, managerial and work procedures, marketing knowledge, reputations and trademarks" (Markusen 1998). Knowledge spillovers are likely to be associated primarily with such knowledge-intensive industries. Within services the sector which currently accounts for the bulk of the global FDI stock the predominant subsectors are finance and business activities. Within manufacturing, the chemical industry is predominant, followed by motor vehicles and electronics. Primary activities account for only a small proportion of global FDI.

In discussing the labor-market effects of FDI, one clearly needs to distinguish between the effects on the host, or destination, economy and those on the home, or source, economy. The following two major subsections follow this distinction.

FDI and the Host-Economy Labor Market How might FDI affect the labor market of the host economy? Obviously, in the presence of unemployment, an FDI inflow might simply lead to new job creation, without further complications. If the inflow is large enough relative to the size of the economy, it might also bid up wages and improve the economy's terms of trade, generating gains even in the absence of market failures. For a small economy like Ireland, which had an unemployment rate of 17 percent before the substantial increase in export-platform FDI of the "Celtic Tiger" era of the 1990s and beyond, both of these considerations apply.

In general however, the matter is less straightforward, and indeed, cross-country growth regressions that attempt to isolate the effects of inward FDI yield inconclusive results. One problem with these studies is that they fail to distinguish between the different characteristics of FDI projects associated with import substituting and outward-oriented regimes. Case study investigations of FDI in the automotive and high-performance electronics industries in developing economies show that the effects depend very much on whether local subsidiaries are integrated into the parent firm's international sourcing networks (as under outward orientation) or whether they are oriented toward protected domestic markets. Foreign direct investment projects of the first type motivated by "global sourcing" considerations have been found to differ from "tariff-jumping" FDI projects of the second type in terms of the size of the local plant, its proximity to the technological frontier, the rapidity with which technology and quality control procedures are upgraded, and the general efficiency of the operations. In each case, the tariff-jumping FDI projects exhibit less desirable characteristics. Nor do they represent an effective infant-industry strategy, while many prominent domestic component producers in outward-oriented

regimes are known to have originated as contract manufacturers to foreign plants (Moran, Graham, and Blomström 2005).

The cross-country studies on FDI also typically do not specify the precise microeconomic channels through which FDI might be thought to influence growth. Besides the direct first-round employment effects of an FDI project, second-round effects will arise through the impact of FDI on local firms, while third-round effects will arise through the impact on public-goods provision and expenditure of any increased tax revenues that might be associated with FDI.

MNE entry can crowd out local firms in the product market, though this outcome is less likely with export-platform FDI, while benefits can arise through the erosion of local monopoly power and the inducement to reduce internal inefficiencies. The investment might also raise the productivity of local firms through knowledge spillovers, deeper markets in specialized factors, and increased scope for backward and forward linkages between customer and supplier firms   the classic "Marshallian externalities."

In the labor market itself, while the short-term value of an FDI project will depend on the extent to which other jobs are crowded out, productivity growth and hence wage growth over the longer term may be higher in multinational firms. The characteristics of the jobs created are also important. If some interindustry wage differences are associated with industry rather than worker characteristics (as in efficiency-wage models, for example), an FDI-induced change in industrial structure can yield benefits in terms of increased wages. Labor-market volatility might also increase, however, because of the ease with which MNEs can switch activities across plants. The latter possibility, however, might also serve to make local labor markets more competitive.

In general, the second-round labor-market effects arise either because MNEs differ in a number of key aspects from the domestic firms they may replace or because FDI induces changes in the behavior of domestic firms.

All the international evidence suggests that MNEs on average are more productive than domestic firms — in terms of both output and value added per employee — and that they exhibit higher productivity growth. These findings are largely accounted for by the fact that MNEs are larger (in terms of output, employment, and value added), they invest more, and they use more intermediate inputs per employee than purely domestic firms. When these differences are controlled for, the productivity gap largely disappears, though the fact that some small element remains suggests that multinationality per se is associated with somewhat greater efficiency in terms of input usage. From a policy perspective, however, the fact that MNEs differ from local firms in these key characteristics is important in that it means that they bring to the host country a set of characteristics and techniques that would otherwise be difficult to establish locally.

The literature is virtually unanimous in its finding — from both developed countries such as the United States, the United Kingdom, and Ireland, and developing countries such as Côte d'Ivoire, Morocco, Venezuela, and Indonesia — that foreign firms pay higher wages than privately owned domestic firms (see, e.g., Aitken, Harrison, and Lipsey 1996; Harrison 1996; Scheve and Slaughter 2005). Part of the reason is undoubtedly that foreign MNEs employ more highly skilled labor than domestic firms do. It is difficult to determine whether this difference fully accounts for the wage differential as skills are often imperfectly measurable. There are a number of reasons, however, why MNEs might pay higher wages than domestic firms for equivalently skilled workers. MNEs may want to minimize the risk that their proprietary knowledge is dissipated through frequent labor turnover or they may wish to avoid excessive geographical disparities. Higher wages might also be required to compensate for perceived disadvantages such as greater employment volatility. Foreign firms also generally provide higher levels of training than domestic firms, resulting in higher subsequent wage growth.

Compared to the other topics discussed, little research has been carried out on the question of the relative volatility of employment in MNEs and domestic firms. Contrary to the presumption that MNEs might be more willing to switch activities across locations in response to wage pressures, the available research suggests that MNE employment levels are less affected by wage changes than is the case for domestic firms. This finding appears likely to arise because of the bundle of characteristics other than ownership per se that distinguishes them from local firms.

A recent study on Irish data adds a new element to the research on volatility. It finds that while the higher-technology MNE-dominated sectors exhibit faster employment growth than the traditional lower-technology sectors populated primarily by domestic firms, employment growth in the former sectors is also more volatile. The application of portfolio theory demonstrates however that the diversification resulting from the attraction of MNEs has allowed the aggregate manufacturing sector to grow more rapidly without a commensurate increase in overall volatility and risk (Barry and Kearney 2006).

We now turn to the question of whether and why FDI might induce changes in the behavior of domestic firms. Multinational and domestic firms interact in a number of ways, some beneficial and some harmful to the latter. The beneficial effects can arise through market transactions or via externalities while the adverse consequences arise through increased competition. The overall effect will determine the second-round employment consequences of foreign direct investment.

The transfer of proprietary assets from MNEs to domestic firms can take place via transactions deliberately negotiated between the parties, as in the case of technology licensing agreements or the provision of direct support for the upgrading of locally provided inputs, which is relatively common in developing countries. Alternatively, technology transfer can take place through externalities that do not bring any direct return to the MNE. Interactions between different firms can generate more knowledge flows than are contracted for. Indeed the inability to write complete contracts is a prime reason

why MNEs frequently prefer to carry out such transactions internally. Even in this case, though, MNE employees who move to local firms will bring with them much of what they have learned in their earlier positions.

Pecuniary as opposed to technological externalities arise when the presence of MNEs leads to increased local investment in increasing-returns activities. There is widespread evidence that MNEs, by making use of local suppliers, have expanded both the quality and the output of the latter. MNE presence can also lead to an expanded provision of public goods such as infrastructure and education, investment in which might not be cost effective unless demand is sufficiently large. Enhanced provision of public goods clearly expands the development potential of the host economy.

MNEs also represent increased competition for domestic firms, however. If market structure is imperfectly competitive, local profit margins are reduced. In increasing-returns sectors, the reduction in market share can force local firms up their average cost curve and away from the minimum efficient scale level of output, possibly causing exit. These negative competitive effects can offset any favorable technological externalities for local firms but are nevertheless likely to be beneficial for the overall economy.

How do these two opposing forces play out in reality? The many empirical studies on the topic do not provide an unambiguous answer to the question. The message that does emerge, however, is that the likelihood of positive second-round employment effects depends on the size of the technological gap between MNEs and domestic firms, the extent of vertical linkages between the two, and the nature of competition in the industries concerned.

Earlier studies based largely on cross-section data almost always found MNE presence to be associated with higher domestic-firm productivity. This correlation might simply be a consequence of the fact that foreign firms tend to invest in higher productivity sectors. Hence more detailed analysis is required.

The importance of the technology gap between foreign and domestic firms is attested to in a number of ways. The gap is likely to be smaller in more developed economies. In line with this, a number of studies find that spillovers are concentrated in middle-income developing countries, with no evidence of such effects in the poorest developing countries. The type of technology and the amount imported by MNEs is also found to be greater in countries and industries where the local labor force is better educated. Beneficial effects on domestic firms are also found to be especially likely to occur when there are vertical linkages between local firms and MNEs, and particularly with upstream local suppliers.

While spillover effects appear strongest in technology-intensive sectors where local firms' capabilities are above a threshold level, competition effects appear to dominate in less technology-intensive sectors. Here, the entry of MNEs tends to lead to an initial decline in the profit margin of local firms, though these can bounce back in the event of competition-induced efficiency gains. A number of studies have also found that MNE presence helps local firms to break into new export markets.

FDI and the Source-Economy Labor Market
Similar questions arise in the case of source-economy labor markets. Here we are interested in the effect of *outward* direct investment (ODI) on employment, production, wages, and the skill mix of the economy, and on the productivity of the firms themselves.

The four channels through which source-economy effects arise are: (1) the redeployment of source-economy labor, if displaced by outward FDI, (2) increased imports of source-economy goods and services by the economies hosting increased FDI, (3) savings accruing to source-economy investors and customers if activities shift to lower-cost locations, and (4) profit repatriation by foreign affiliate companies. All of these will have knock-on effects on the labor market.

The short-term employment effect will depend on whether the overseas and domestic activities of the firm are complements or substitutes. The qualitative employment effects on the home base may differ between the short term and the longer term also, in that an ODI-induced expansion in the market share of the MNE can strengthen home-economy opera-

tions even if home activities appear to be substitutable for the firm's foreign operations in the short run. The counterfactual also should be taken into account, in that even if ODI leads to a contraction in home-location operations, the alternative to ODI may be the complete closure of these operations.

The question of whether home and host-economy activities are substitutes or complements is related to whether FDI is vertical or horizontal in nature. Vertical FDI entails the geographic fragmentation of the production process, while investments undertaken in order to gain an advantage in supplying local or regional markets are classified as horizontal. In general, vertical FDI (typically from developed to outward-oriented developing economies) is associated with complementarity between home and host activities. It might be thought that horizontal FDI (typically between developed economies) will be associated with substitutability as it can displace home exports. Foreign plants however frequently import inputs or other complementary products from the MNE's home base, and studies of U.S., Japanese, Swedish, and other European MNEs have found foreign-affiliate sales to be associated with an increase in the exports of the home-based parent MNE (see, e.g., Lipsey, Ramstetter, and Blomström 2000). Higher capital expenditures on the part of foreign affiliates of U.S. MNEs have also been found to be associated with higher U.S. investments by the parent companies (Desai, Foley, and Hines 2005). The overall implication is that home and foreign production are combined to generate final output at lower cost than would be possible without ODI. Thus production is ultimately increased in both locations. Even if this were not true at the level of the individual firm, labor-market flexibility can serve to ensure that the home economy remains at or close to full employment.

We now turn to the effect of outward FDI on the relative demand for skilled and unskilled labor. If firms in high-income countries fragment their production process and outsource labor-intensive stages to lower-wage countries, this outsourcing leads to a relative increase in the demand for skilled labor at home. Headquarters services are a particularly highly skilled segment, comprising R&D, design, marketing, finance, and strategic management. An increase in demand for such services will show up as an increase in the relative wages of more highly skilled labor, as has been seen in the developed world over recent decades.

The offshoring of unskilled-labor-intensive segments of the production process does entail real welfare losses for displaced workers, however. Evidence from the United States shows that about one-third of displaced workers will not have found a job one year after displacement, though most will have found new jobs after a further lag. Frequently, however, the new jobs found by such workers are at reduced wage levels. The types of workers who suffer most in the structural adjustment process are those of advancing age with low educational attainment, and those who have had a long period of tenure in the lost job (reflecting the depreciated value of their firm-specific and industry-specific skills). There is widespread agreement that public intervention to minimize dislocation resulting from structural change should be focused on providing the displaced with new skills and training and ensuring that they do not become detached from the labor market. While the economy gains in the long run, the adverse adjustment costs will be higher if labor-market rigidities hinder employment growth.

MNEs also establish foreign operations in order to source new technology, which tends to benefit both the companies themselves and the home economy. National productivity has been shown to be increased by outward investments in other R&D-intensive economies, while other research suggests that the international networks that MNEs establish can provide other home-based companies with links — through interaction with these firms — to innovation in other areas of the globe.

The evidence summarized here leads to generally positive conclusions. Inward FDI can yield positive benefits for domestic firms through externalities and other channels, while, to the extent that it causes exit of the least efficient incumbent local firms, this elimination of inefficient firms is itself associated with an increase in productivity and hence in wages.

In the case of outward FDI, foreign-affiliate output and employment are frequently complementary to parent-firm home activities and, even when they are not, ODI tends to boost domestic productivity so that home-economy wages rise even if new employment opportunities are located outside the original firm.

As in all cases of structural adjustment, labor-market flexibility is important to ensure that new employment opportunities materialize as existing jobs disappear. The benefits of structural adjustment frequently accrue to workers with different characteristics from those displaced in the adjustment process. Hence the discussion on labor-market interventions to deal with displaced workers raises issues similar to those encountered in the policy discussion on trade adjustment assistance (see, e.g., Baicker and Rehavi 2004; Kletzer 2004).

See also agglomeration and foreign direct investment; domestic content requirements; factor endowments and foreign direct investment; footloose production; foreign direct investment and tax revenues; location theory; outsourcing/offshoring; subsidies and financial incentives to foreign direct investment; technology spillovers; unions and foreign direct investment; vertical versus horizontal foreign direct investment

FURTHER READING

Aitken, Brian, Anne Harrison, and Robert Lipsey. 1996. "Wages and Foreign Ownership: A Comparative Study of Mexico, Venezuela, and the United States." *Journal of International Economics* 40 (3 4): 345 71. Shows that for Mexico and Venezuela, unlike in the case of the United States, there is no evidence of wage spillovers from foreign investment leading to higher wages for domestic firms.

Baicker, Katherine, and M. Marit Rehavi. 2004. "Trade Adjustment Assistance." *Journal of Economic Perspectives* 18 (2): 239 55. Presents a history of trade adjustment assistance and a discussion of the policy issues.

Barry, Frank, and Colm Kearney. 2006. "MNEs and Industrial Structure in Host Countries: A Portfolio Analysis of Irish Manufacturing." *Journal of International Business Studies* 37 (3): 392 406. Employs portfolio theory to analyze the characteristics of Ireland's stock of industries.

Desai, Mihir, C. Fritz Foley, and James R. Hines Jr. 2005. "Foreign Direct Investment and the Domestic Capital Stock." *American Economic Review* 95 (2): 33 38. Evaluates evidence of the impact of outbound foreign direct investment on U.S. domestic investment rates.

Görg, Holger, and David Greenaway. 2004. "Much Ado about Nothing? Do Domestic Firms Really Benefit from Foreign Direct Investment?" *World Bank Research Observer* 19 (2): 171 97. Survey of the literature on FDI spillovers.

Harrison, Anne. 1996. "Determinants and Effects of Direct Foreign Investment in Côte d'Ivoire, Morocco, and Venezuela." In *Industrial Evolution in Developing Countries*, edited by M. J. Roberts and J. R. Tybout. New York: Oxford University Press for the World Bank, 163 86. Côte d'Ivoire here emerges as an exception to the general finding that foreign firms pay higher wages than domestic firms.

Kletzer, Lori G. 2004. "Trade related Job Loss and Wage Insurance: A Synthetic Review." *Review of International Economics* 12 (5): 724 48. Develops a set of stylized facts of trade related job loss, with a focus on worker characteristics and labor market consequences.

Lipsey, Robert. 2004. "Home and Host Country Effects of Foreign Direct Investment." In *Challenges to Globalization: Analyzing the Economics*, edited by Robert E. Baldwin and L. Alan Winters. Chicago: University of Chicago Press, 333 79. Broad survey of the literature.

Lipsey, Robert, Eric Ramstetter, and Magnus Blomström. 2000. "Outward FDI and Parent Exports and Employment: Japan, the United States, and Sweden." *Global Economic Quarterly* 1 (4): 285 302. Studies the relationship between foreign affiliate sales and parent company exports and employment for Japanese, U.S., and Swedish multinational firms.

Markusen, James. 1998. "Multinational Firms, Location and Trade." *World Economy* 21 (6): 733 56. Sets out the "knowledge capital" model of the multinational firm.

Moran, Theodore H., Edward M. Graham, and Magnus Blomström, eds. 2005. *Does Foreign Direct Investment Promote Development?* Washington, DC: Institute for International Economics. Includes discussion of the

differences between FDI in protectionist and outward oriented regimes.

Navaretti, Giorgio Barba, and Anthony Venables. 2004. *Multinational Firms in the World Economy*. Princeton, NJ: Princeton University Press. A summary of theory and evidence on FDI and multinational enterprises.

Scheve, Kenneth F., and Matthew J. Slaughter. 2005. "Foreign Direct Investment and Labour Market Out comes." In *The Internationalisation of Asset Ownership in Europe*, edited by Harry Huizinga and Lars Jonung. Cambridge: Cambridge University Press, 77 109. Broad survey of the literature.

FRANK BARRY

■ foreign direct investment and tax revenues

Does foreign direct investment (FDI) expand tax revenues? It is commonly believed that FDI enhances economic growth, both directly by increasing the stock of physical capital in the recipient economy, and indirectly by inducing human capital development and promoting technological upgrading. Based on evidence of positive effects of FDI on economic growth, international trade, and domestic investment (under certain conditions), it is then claimed that FDI also expands government tax revenues. The reasoning is that higher levels of economic activity generate higher levels of production, and consequently higher levels of government tax revenue, all else being equal. This perceived additional benefit of greater tax revenues has led host countries to expand their attraction packages to include firm-specific incentives such as tax holidays, tax breaks, and subsidies.

These potential benefits have given multinational corporations (MNCs) a strong bargaining position to negotiate favorable conditions for their entry into new markets. As a result, MNCs are often able to extract additional concessions in the negotiation process, in some cases even long-lived tax breaks and exemptions. Even if FDI has the potential to expand government tax revenues by enhancing economic growth, the fact that MNCs receive such tax incentives creates distortions that can contradict the main goals of tax policy. It can also affect the competitiveness of domestic firms and, consequently, influence the collection of tax revenues. Therefore, a direct link between FDI and economic growth does not necessarily translate directly to expanded government tax revenues.

The performance of government tax revenues is determined by three main aspects: *structural characteristics* (factors that influence the tax handles of an economy), such as income level, population size, the existence of informal or subsistence sectors, and the dependence on taxation of trade; *macroeconomic characteristics*, such as the governmental debt stance, the government's current and past budget position, and the degree of reliance on inflationary tax to generate revenue; and *external characteristics*, such as the terms of trade, exchange rate volatility, and the degree of trade liberalization of the economy. Changes in one or more of these factors can affect the performance of tax revenues; given that FDI can affect some of these factors, it can also potentially affect the collection of tax revenues. FDI's effect on government tax revenues in the host country may also differ from the effect on the source country.

FDI's Effect on the Source Country's Tax Revenues Even if FDI directs investment into foreign countries, the parent enterprises still maintain headquarter activities in the source country and thus generate tax responsibilities that affect the collection of tax revenues. Although FDI generates economic growth and spillovers in the countries that attract it, it mainly benefits the owners of the MNCs. Subsidiaries repatriate profits to their respective parent companies through dividends, interest, and royalty payments, which are taxed by the source country. In addition, most research and development is conducted in the source country, which generates significant taxable economic activity in the country of origin, thus again affecting the collection of tax revenues.

Most countries also offer relief from double taxation to MNCs participating in outward foreign investment. Some countries allow foreign tax credits to be applied to domestic tax liabilities, deductions of

taxes paid to foreign governments from their tax liabilities, exemptions for the foreign earnings of their MNCs from domestic taxation, or some combination of the three. Although this relief from double taxation tends to lower the collection of tax revenues from production abroad, it protects the health of the MNCs of the host country.

Furthermore, since the performance of the MNCs affects the citizens of the source country   the owners of the firm   and MNCs bear the risk of expropriation that is inherent in foreign investment, there is an incentive for both the government and MNCs to protect and enhance the revenues from FDI through bilateral treaties. These treaties lower trade barriers, protect foreign investment, reduce tax rates abroad, and reduce tax uncertainty over the long term (Chisik and Davies 2004).

FDI's Effect on the Host Country's Tax Revenues FDI inflows are expected to fuel economic growth and generate higher levels of output in the host country. Conventional wisdom presumes that FDI would consequently increase government tax revenues. FDI can affect government tax revenues in a variety of ways: it can directly increase domestic taxes on income and goods and services if the higher levels of foreign investment increase production and employment; it can indirectly increase (1) taxes on international trade if the product is sold abroad or if imported inputs are used in the production process, (2) taxes on income and goods and services if it fosters greater activity in domestic firms participating in the production chain, and (3) overall taxes if higher or better incomes emanating from the previous points find their way back into the economy (multiplier effect); or it can affect the collection of tax revenues through the formalization of economic activity brought about by the existence or addition of foreign investment.

Although such considerations could have a positive impact on the collection of tax revenues, the intensive use of tax incentives to attract FDI may introduce potential distortion to this link, which can belie the conventional wisdom. In addition, although the greater economic activity brought about by FDI

should raise tax revenues in the long run, tax incentives tend to reduce tax revenues in the short run.

Since FDI goes hand in hand with trade liberalization   MNCs account for two-thirds of all cross-border sales   it is this area that has received the most attention. Trade tax revenues usually decrease with trade liberalization, and liberalization's effect on overall taxes depends on the efficiency of the transition toward domestic taxes (on goods and services), the resulting volatility of the exchange rate (and its effect on import and export bases), and the adjustment of exports and imports due to changes in the terms of trade.

FDI can also affect government tax revenues through formalization of the economy. Since greater foreign investment increases production in the formal sector, it also formalizes the production of domestic firms supplying inputs to the MNC, thus contributing to the collection of tax revenues. It also strengthens tax compliance of domestic companies by exposing them to the best business practices and corporate governance of the MNCs.

Although the potential impact of FDI on government tax revenues is considerable, empirical research in this area is almost nonexistent, with the exception of the work of Braunstein and Epstein (2004) and Vacaflores (2006). The former find that inward FDI has a negative impact on tax revenue in Chinese provinces   which is ameliorated by its positive impact on trade   while the latter shows that FDI has a positive impact on total tax revenues in Latin America, which is channeled through its impact on the main component, taxes on goods and services.

Despite this lack of conclusive empirical results, host countries are increasingly complaining about the tax evasion that tax treaties sometimes allow, the detrimental tax breaks that they were forced to give in the capitalization/privatization process, and the manipulation of intrafirm prices, all of which are perceived as detrimental to the tax collection of host countries.

Policy Implications The anticipation that a country's tax structure affects the MNC's choice to locate in a particular country has led most countries to adjust their tax system, making it more flexible to

international firms, becoming in many cases even more attractive for MNCs than domestic firms. This subsidization of FDI through tax incentives brings to light the potential harm that greater foreign participation can inflict on domestic firms, as the partial or total crowding out of domestic firms increases the possibility that some domestic factors of production may lose with higher levels of FDI. Glass and Saggi (1999) show that if the impact of MNCs on the profitability of domestic firms is sufficiently negative, FDI may lower host-country welfare to a degree that the optimal policy toward FDI should instead be a higher level of taxation. So the degree of substitutability of production between domestic and foreign firms and/or complementarities in their production processes become important in the overall collection of tax revenues.

To better understand the overall effects of FDI, more needs to be known about the effects of FDI on government tax revenues. Theoretical studies suggest that this effect should be positive, at least in the long run once tax incentives have expired. However, the inherent adjustments in tax systems and potential impact on the performance of domestic investment require further research to establish the direction and magnitude of such effect more conclusively.

See also agglomeration and foreign direct investment; footloose production; foreign direct investment and labor markets; foreign direct investment under monopolistic competition; foreign direct investment under oligopoly; international investment agreements; intrafirm trade; location theory; trade-related investment measures (TRIMs); transfer pricing

FURTHER READING

Braunstein, Elissa, and Gerald Epstein. 2004. "Bargaining Power and Foreign Direct Investment in China: Can 1.3 Billion Consumers Tame the Multinationals?" In *Labor and the Globalization of Production: Causes and Consequences of Industrial Upgrading*, edited by Will Milberg. London: Palgrave Macmillan. An empirical analysis of the effects of FDI on economic growth, employment, investment, and tax revenues for Chinese provinces.

Chisik, Richard, and Richard B. Davies. 2004. "Gradualism and Tax Treaties with Irreversible Foreign Direct Investment." *International Economic Review* 45 (1): 113 39. Provides a concise description of tax treaties and their influence on tax rates for FDI.

Glass, Amy, and K. Saggi. 1999. "FDI Policies under Shared Factor Markets." *Journal of International Economics* 49 (2): 309 32. A study of the impact of MNCs (FDI) on the profitability of domestic firms, host country welfare, and optimal tax policy.

Hanson, Gordon H. 2001. "Should Countries Promote Foreign Direct Investment?" United Nations Conference on Trade and Development, G 24 Discussion Paper Series No 9. Geneva: UNCTAD. A useful analysis of the impact of FDI on national welfare, with a practical summary of policies used to promote FDI.

Vacaflores, Diego E. 2006. "The Effect of Foreign Direct Investment on Tax Revenues in Latin America: Are the Facts in Accord with the Conventional Wisdom?" Texas State University Working Paper. San Marcos, TX. Examines the effect of FDI on tax revenue performance for a group of Latin American countries, for total tax revenues and its main components.

DIEGO E. VACAFLORES

foreign direct investment: the OLI framework

The OLI, or "eclectic," approach to the study of foreign direct investment (FDI), developed by John Dunning, has proved to be an extremely fruitful way of thinking about multinational enterprises (MNEs) and has inspired a great deal of applied work in economics and international business. In itself it does not constitute a formal theory that can be confronted with data in a scientific way, but it nevertheless provides a helpful framework for categorizing much (though not all) recent analytical and empirical research on FDI. This survey first summarizes the OLI paradigm and then uses it as a lens through which to review some of the highlights of this research, while also noting some important issues that it neglects.

OLI stands for ownership, location, and internalization, three potential sources of advantage that may underlie a firm's decision to become a multinational. Ownership advantages address the question of why some firms but not others go abroad, and suggest that a successful MNE has some firm-specific advantages that allow it to overcome the costs of operating in a foreign country. Location advantages focus on the question of where an MNE chooses to locate. Finally, internalization advantages influence how a firm chooses to operate in a foreign country, trading off the savings in transactions, hold-up, and monitoring costs of a wholly owned subsidiary against the advantages of other entry modes such as exports, licensing, or joint venture. A key feature of this approach is that it focuses on the incentives facing individual firms. This is now standard in mainstream international trade theory but was not at all so in the 1970s, when FDI was typically seen through a Heckscher-Ohlin lens as an international movement of physical capital in search of higher returns (see, for example, Mundell 1957).

Ownership Ownership advantages are key to explaining the existence of MNEs. A central idea is that firms are collections of assets, and that candidate MNEs possess higher-than-average levels of assets having the character of internal public goods. These assets can be applied to production at different locations without reducing their effectiveness. Examples include product development, managerial structures, patents, and marketing skills, all of which are encompassed by Helpman's (1984) catchall term *headquarter services*. Although this is clearly a multidimensional factor, it is common to model it in terms of a single index of firm productivity. The most sophisticated treatment along these lines is found in work on heterogeneous firms by Helpman, Melitz, and Yeaple (2004), which combines the simplest version of the horizontal motive for FDI (to be discussed later in this article) with the assumption that firms differ in their productivities. A potential firm must pay a sunk cost to determine its productivity, and when this is revealed, active firms sort themselves into different modes of production. Low-productivity firms produce only for the home mar-

ket; medium-productivity firms choose to pay the fixed costs of exporting; but only the most productive firms choose to pay the higher fixed costs of engaging in FDI. These predictions are consistent with the evidence. As a further contribution, the paper derives from the model the prediction that industries with greater firm heterogeneity will have relatively more firms engaged in FDI and shows that this prediction is confirmed by the data. This work and others like it do not explore, however, why firm productivities differ in the first place. Prior investment in research and development (both process and product) and in marketing presumably account for the disproportionately greater productivity of most MNEs.

Location While international trade theory has tended to take ownership advantages for granted or else to model them in fairly obvious ways, rather more attention has been devoted to exploring alternative motives for MNEs to locate abroad. A key issue that has attracted much attention is the distinction between "horizontal" and "vertical" FDI. Horizontal FDI occurs when a firm locates a plant abroad in order to improve its market access to foreign consumers. In its purest form, this simply replicates its domestic production facilities at a foreign location. Vertical FDI, by contrast, is not primarily or even necessarily aimed at production for sale in the foreign market, but rather seeks to take advantage of lower production costs there. Since in almost all cases the parent firm retains its headquarters in the home country, and the firm-specific or ownership advantages can be seen as generating a flow of headquarter services to the host-country plant, there is a sense in which all FDI is vertical. Nevertheless, the distinction between market-access and cost motives for FDI is an important one.

The horizontal motive for FDI reflects what Brainard (1997) has called a "proximity-concentration trade-off": building a local plant saves on trade costs and so has the advantage of proximity; but it loses the benefits of concentrating production in the firm's home plant. Let $\pi^*(t^*)$ denote the operating profits that a potential MNE can earn from selling in a foreign market subject to per unit trade costs t^* (which can include both tariffs and transportation

costs). These operating profits are decreasing in t^*: higher trade costs reduce operating profits. Constructing a local plant avoids the trade costs, leading to higher operating profits of $\pi^*(0)$; however, it requires an additional fixed cost f. Hence the trade-cost-jumping gain, the difference between the total profits from FDI, Π^F, and those from exporting, Π^X, equals:

$$\gamma(\underset{+}{t^*}, f) \equiv \Pi^F - \Pi^X = \pi^*(0) - f - \pi^*(t^*) \quad (1)$$

Thus FDI is encouraged relative to exports by proximity (lower trade costs t^*) but discouraged by the benefits of concentration (higher fixed costs f).

The vertical motive for FDI implies a very different view of the determinants and implications of FDI. Now the focus is on how a firm can serve its home market: either by producing at home, or by vertically disintegrating and moving its production facilities to a cheaper foreign location. Assuming for simplicity that each unit of output requires a single unit of labor, we can write the operating profits of serving the home-country market as $\pi(c)$, where c includes both factor costs and trade costs. If the firm remains a domestic firm and supplies its home market from its parent plant, where w is the local wage rate, it incurs no trade costs so its profits Π^D will equal $\pi(w)$. Alternatively, it can engage in FDI and locate a new plant in the host country, exporting all its output back to the source country and incurring a trade cost of t. In that case, it incurs a plant-specific fixed cost f as in the case of horizontal FDI, and earns operating profits of $\pi(w^* + t)$, where w^* is the host-country wage. The relative profitability of FDI is therefore:

$$\Pi^F - \Pi^D = \mu(w^* + t, w) - f$$
$$\text{where:} \quad \mu(w^* + t, \underset{+}{w})$$
$$\equiv \pi(w^* + t) - \pi(w) \quad (2)$$

Now the decision to engage in FDI depends on the trade-off between the benefits of concentration, on the one hand, and the cost savings from offshoring, on the other, where the latter are denoted by the term $\mu(w^* + t, w)$. This offshoring gain depends negatively on the host-country wage w^* and positively on the source-country wage w: the vertical

motive for FDI attaches great importance to comparative costs of production. In addition, the gain is decreasing in the source-country trade costs t, implying plausibly that trade liberalization will encourage FDI.

Empirical studies of FDI have until recently tended to favor the horizontal over the vertical motive. For example, many case studies have shown that "tariff-jumping" has been important in many historical episodes. It has also been noted that the bulk of FDI is between high-income countries with relatively similar wage costs (though much of this is likely to be neither vertical nor horizontal FDI, but rather cross-border mergers and acquisitions, to be discussed further later). More formal econometric studies have shown that the horizontal motive provides a good explanation for FDI (see, for example, Brainard 1997; Markusen 2002). On the other hand, there is no clear evidence that FDI falls in importance with distance, as the horizontal model implies. In addition, more recent empirical work by Yeaple (2003b) and others, based on data at the level of individual firms, suggests that both motives are important. It is easy to see why this might be so even in the simple two-country case discussed above. If the foreign market is sizable, then the total gain from FDI as opposed to producing at home (in each case serving both domestic and foreign customers from a single plant) is given by the sum of (1) and (2) above: both trade-cost-jumping and offshoring gains have to be taken into account. More generally, with many countries there are additional reasons for FDI, and the two motives are likely to interact in complicated ways. For example, even for vertically integrated firms, proximity and concentration are not in conflict where serving a group of foreign countries is concerned. The reduction of trade costs between European countries in the 1990s encouraged American and Asian firms serving European markets to concentrate their production in European plants and so engage in "export-platform" FDI. Similarly, Yeaple (2003a) has shown that the horizontal and vertical motives may reinforce each other if a parent firm wishes both to serve foreign markets in similar high-income countries and to avail itself of lower

production costs in low-income countries. In general, therefore, the pattern of location of foreign plants is likely to reflect the complex integration strategies of firms having both vertical and horizontal motives for engaging in FDI.

Internalization Internalization, the third strand of Dunning's taxonomy, is often seen as the most important; in the words of Ethier (1986), "Internalization appears to be emerging as the Caesar of the OLI triumvirate." Explaining why some activities are carried on within firms and others through arm's-length transactions is a major research topic for microeconomics as a whole, not just for the economics of FDI. A pioneering 1937 paper by Ronald Coase argued that the optimal scale of the firm, or the optimal degree of internalization, reflects a balance between the transactions costs of using the market and the organizational costs of running a firm. In recent decades, economists working in information economics have tried to endogenize these two sources of costs, emphasizing the inability of agents to write complete contracts. An early application of this approach to FDI was by Ethier (1986). In his model production requires prior research, the results of which can either be carried out within a vertically integrated firm (in the MNE case) or sold to downstream users. The end user must agree to purchase the research before its outcome is known, however. Ethier shows that a greater degree of uncertainty about the likely success of research efforts makes it more costly for the upstream and downstream firms to write a contract, which because of the complexity of the research process must necessarily be independent of the outcome. Hence more uncertainty raises the likelihood that production will be vertically integrated through MNEs. Moreover, the emergence of MNEs does not require international differences in factor prices, unlike other models of vertical FDI.

A different approach to endogenizing the internalization decision, though also relying on incomplete contracts, is taken by Antras and Helpman (2004). Following the Grossman-Hart-Moore property-rights approach to the problem of bargaining between a firm owner and a potential supplier/employee, ex post efficiency is greater when residual ownership rights are allocated to the party that contributes more to the final output. Embedded in a model of product differentiation and trade, this implies that more efficient firms and firms for which headquarter services are more important should exhibit internalization (the owner contracts with the supplier, who becomes an employee), while less efficient firms should exhibit arm's-length trade (the supplier remains a separate legal entity). In addition the model assumes that final-goods producers are located only in one country, the North of a two-country North-South model. Such producers are assumed to have a twofold choice: on the one hand, they have to choose between vertical integration, which solves the hold-up problem but at the cost of reducing incentives to the provider of the input, and an arm's-length relationship; on the other hand, they could locate their production in either country, trading off higher wages in the North against lower contract protection in the South. The full range of potential outcomes is shown in table 1, and Antras and Helpman show how heterogeneous firms will sort into these different modes, based on their productivity, on the share of headquarter services in the value of output, and on the differences in costs between home and foreign locations.

Cross-Border Mergers Both the OLI framework and the bulk of academic work on FDI until very recently have concentrated on the "greenfield" mode of FDI, where the parent firm constructs a new plant in the host country. Yet in reality the bulk of FDI, especially between developed countries, takes the form of cross-border mergers and acquisitions (M&As), in which the parent firm acquires a con-

Table 1
Taxonomy of location internalization modes

	Location	
	Home	Abroad
Internal	Integrated national firm	FDI
External	Outsourcing	Offshoring

trolling interest in an existing host-country firm. (UNCTAD estimates suggest that M&As accounted for more than 80 percent of worldwide FDI in the 1990s.) The distinction matters, since recent research suggests that the determinants and implications of cross-border M&As are very different from those of greenfield FDI.

Domestic M&As have been extensively studied by scholars in finance and industrial organization, and these literatures suggest two principal motives for them. A "synergy" motive arises in any market where an acquired firm has assets that are complementary to those of the acquirer, whereas a "strategic" motive arises in oligopolistic markets (i.e., markets in which the number of competitors is small), since a firm gains from absorbing a rival and so increasing its own market power.

Postmerger synergies can arise from many sources, including cost savings via internal technology transfer, reductions of overhead and other fixed costs, and the integration of pricing and marketing decisions on differentiated products. In an open-economy context, a particularly plausible kind of synergy is between the "O" and "L" advantages of different firms: the superior productivity and international networks of an acquiring MNE, on the one hand, combined with the local knowledge and distribution network of a potential target firm, on the other. Nocke and Yeaple (2008) develop a model that captures this kind of synergy: a competitive international market for corporate assets allows firms to match with suitable affiliates. Their model predicts that efficient matching occurs: more efficient parent firms acquire more efficient targets. They also show, however, that the most efficient firms engage in greenfield FDI rather than in cross-border M&As, a result consistent with the evidence.

Mergers driven by synergies may be expected to raise world welfare, provided the synergies are realized in practice. By contrast, mergers driven by strategic considerations might be expected to reduce welfare since they increase concentration. Neary (2007) shows, however, that this intuition is incomplete for two reasons. First, in the absence of synergies, the only mergers that will occur in equi-

librium are those in which the acquirer can afford to buy out the target firm. This implies that the target firm must be considerably smaller, and therefore eliminating it is likely to enhance global efficiency. Second, in general equilibrium, the expansion of more efficient acquiring firms and the elimination of less efficient target firms puts downward pressure on wages, thus encouraging increased output and lower prices in all sectors. Hence mergers are likely to raise overall welfare, although income distribution shifts in favor of profits at the expense of wages. Neary's model also makes the positive prediction that mergers take place in the same direction as trade and therefore they are encouraged rather than (as in the horizontal model of greenfield FDI) discouraged by decreases in trade costs. In line with empirical evidence, cross-border mergers thus serve as "instruments of comparative advantage," encouraging more specialization and trade along comparative advantage lines.

The OLI Framework: An Overall Assessment
In conclusion, the OLI framework does not directly address one of the key issues that has dominated economists' thinking about FDI, the distinction between horizontal and vertical motives for locating production facilities in foreign countries. Nor does it address the increasingly important distinction between greenfield and M&A modes of engaging in FDI. Nevertheless it remains a helpful way of organizing thinking about one of the most important features of the world economy.

See also internalization theory; location theory; multinational enterprises

FURTHER READING
Antras, Pol, and Elhanan Helpman. 2004. "Global Sourcing." *Journal of Political Economy* 112 (3): 552 80. A pioneering model that shows how firms with different productivities will choose between locating production at home or abroad and keeping it within the firm or outsourcing it to a subcontractor.

Barba Navaretti, Giorgio, Anthony J. Venables, et al. 2004. *Multinational Firms in the World Economy.* Princeton, NJ: Princeton University Press. An invaluable

overview of theoretical and empirical work on greenfield FDI.

Brainard, S. Lael. 1997. "An Empirical Assessment of the Proximity Concentration Tradeoff between Multi national Sales and Trade." *American Economic Review* 87 (4): 520 44. A landmark paper in the empirical appli cation of the horizontal model of FDI.

Dunning, John H. 1977. "Trade, Location of Economic Activity, and the MNE: A Search for an Eclectic Ap proach." In *The International Allocation of Economic Activity*, edited by Bertil Ohlin, Per Ove Hesselborn, and Per Magnus Wijkman. London: Macmillan, 395 418. The first statement of the OLI approach, later re fined and extended in many books and papers by the author and his collaborators.

Ethier, Wilfred J. 1986. "The Multinational Firm." *Quar terly Journal of Economics* 101 (4): 805 34. An im portant paper that models the internalization deci sion in a general equilibrium model with incomplete contracts.

Helpman, Elhanan. 1984. "A Simple Theory of Interna tional Trade with Multinational Corporations." *Journal of Political Economy* 92 (3): 451 71. A pioneering ex ploration of vertical FDI, embedded in a Heckscher Ohlin model with monopolistic competition.

Helpman, Elhanan, Marc Melitz, and Stephen Yeaple. 2004. "Exports versus FDI with Heterogeneous Firms." *American Economic Review* 94 (1): 300 16. Extends the horizontal model of FDI to allow for productivity dif ferences across firms.

Markusen, James R. 2002. *Multinational Firms and the Theory of International Trade*. Cambridge, MA: MIT Press. An invaluable overview of the author's contribu tions, both alone and with coauthors, especially Horst mann and Venables. Emphasizes horizontal greenfield FDI.

Mundell, Robert. 1957. "International Trade and Factor Mobility." *American Economic Review* 47 (3): 321 35. A key reference in the Heckscher Ohlin inspired literature on FDI. An important step in the elaboration of the factor endowments approach to trade flows. Through its emphasis on trade costs, it was a precursor of the hori zontal view of FDI. However, its focus on sectors rather than firms, and its counterfactual prediction that FDI is mainly driven by international differences in rates of return, stand in contrast with the OLI approach, and have been abandoned in more recent work.

Neary, J. Peter. 2007. "Cross border Mergers as Instru ments of Comparative Advantage." *Review of Economic Studies* 74 (4): 1229 57. A model of cross border mergers in general oligopolistic equilibrium, focusing on strategic motives.

———. 2008. "Trade Costs and Foreign Direct Invest ment." In *Foreign Direct Investment and the Multi national Enterprise*, edited by S. Brakman and H. Garretsen. Cambridge, MA: MIT Press, 13 38. An overview of recent work, highlighting that export platform FDI and cross border mergers can overcome the counterfactual prediction of the simple horizontal model that FDI should fall as trade costs fall.

Nocke, Volker, and Stephen Yeaple. 2008. "An Assignment Theory of Foreign Direct Investment." *Review of Eco nomic Studies* 75 (2): 529 57. A model with monopo listically competitive firms, highlighting the synergy motive for cross border mergers and acquisitions.

Yeaple, Stephen. 2003a. "The Complex Integration Strat egies of Multinational Firms and Cross Country Dependencies in the Structure of Foreign Direct In vestment." *Journal of International Economics* 60 (2): 293 314. A model showing that vertical and horizontal motives can reinforce each other in a multicountry world.

———. 2003b. "The Role of Skill Endowments in the Structure of U.S. Outward Foreign Direct Investment." *Review of Economics and Statistics* 85 (3): 726 34. A careful empirical study using firm level data which shows that U.S. outward FDI is driven by both vertical and horizontal influences.

J. PETER NEARY

■ foreign direct investment under monopolistic competition

Central questions in the study of foreign direct in vestment (FDI) are: Where do firms choose to locate their production? To what extent do firms become multinational and supply foreign markets through production by local subsidiaries rather than through trade? What are the associated levels of FDI and

patterns of trade? Theories of FDI under monopolistic competition provide a framework to address these questions. They analyze firms' choices on where to locate production, and also endogenize the number of active firms operating in each country.

Monopolistic competition models typically focus on horizontal rather than vertical FDI. That is, they look at situations where FDI involves producing the same good (or at least undertaking the same stage of production) in several locations. This is in contrast to vertical FDI, where FDI takes the form of offshoring different parts of the production process, although a monopolistic competition framework has also been used to address this issue (for example, Helpman and Krugman 1985).

Theory There are three main elements in a model of monopolistic competition and FDI: equilibrium of the firm, equilibrium of the industry, and general equilibrium of the economy. This review focuses on the first two of these. Each firm has increasing returns to scale, typically modeled by two components: increasing returns at the level of the firm and increasing returns at the level of the plant. At the firm level, increasing returns may derive from the costs of headquarters operations or research and development (R&D); these can be represented by fixed cost F. Plant-level operations have increasing returns to scale, modeled by fixed costs associated with setting up a plant, P. Unit operating costs are typically assumed constant (equal to marginal costs), but contain two different components, production costs and costs of trading internationally, such as transportation costs or trade taxes.

The trade-off faced by firms is clear. If they supply all markets from a single country (and single plant) they save the fixed costs of setting up a new plant but incur additional trade costs. Conversely, FDI means paying for a new plant, but saving trade costs. This is sometimes referred to as the proximity-concentration trade-off. Firms' decisions depend on the levels of these various costs and since fixed costs are involved also on expected sales volumes.

Turning to the industry equilibrium, there is free entry and exit of firms, so the equilibrium number of firms in each country is determined by zero profit

conditions. Competition is imperfect and firms set a markup of price over marginal cost, thus covering their fixed costs. Imperfect competition is typically modeled via product differentiation, often using the Dixit-Stiglitz (1977) framework so that price-cost markups are constant. In the simplest cases all firms in the industry have identical cost and demand functions (e.g., Markusen and Venables 2000).

What equilibrium outcomes are possible? Suppose first that there are two countries of similar size and with similar factor prices. Then the two key parameters are trade costs and the plant fixed cost relative to the firm-level fixed cost, (P/F). If trade costs are low relative to P/F then each firm will operate in a single country (with just one plant) and export to the other country. There will be no FDI, but there will be large volumes of intraindustry trade. If the two countries are identical, intraindustry trade will be balanced; otherwise the larger country will be the net exporter of output from this sector, since trade costs tend to cause firms to locate in the larger market.

If trade costs are high relative to plant fixed costs (or to the ratio P/F), then firms substitute FDI for exports. In the case where the two countries are identical, there is two-way FDI, and each market is supplied by domestic firms and foreign affiliates. There is no trade in goods, although there is trade in services, in the sense that affiliates' earnings cover some of the firm-level fixed costs and can be viewed as payment for headquarters' services. If the two countries are of different sizes, equilibrium has the following pattern: the smaller country may be supplied only by multinationals, while the larger is supplied by both multinational firms and "national" firms producing in a single country. In a multi-country model country size should be interpreted to include the market access of each country. For example, a firm may use a small country as an export platform to supply a larger neighborhood (Ireland and the EU being the obvious example).

In the model sketched earlier, all firms based in one country have the same behavior, this property following from the assumption that they all have the same technologies and face the same demand functions. This has the unattractive consequence that no

country will both export and have outward FDI in the same product. There are a number of ways of generalizing the model to get away from this stark result. One is to make firms within each country heterogeneous, and this approach has been pursued by Helpman, Melitz, and Yeaple (2004). All firms in a particular country face the same fixed costs, but there are now four such fixed costs incurred: by entry, by commencing production, by exporting, and by undertaking FDI. When the firm enters it learns its unit production cost (drawn from a distribution, and therefore varying across firms), and trade costs are a constant multiple of this. Each firm has to make three decisions. The first is whether to produce at all, or to exit; the second is whether or not to supply the foreign market; the third is whether to do so by exports or by FDI. Each of these decisions depends, in a natural way, on the level of the fixed costs associated with each activity relative to the per-unit cost.

The industry equilibrium, with free entry of firms, may now involve some firms exporting and others engaging in FDI. If there are two identical economies, then the following outcomes arise. Firms that commence production but have relatively high unit costs will produce only for the local market, neither exporting nor undertaking FDI. To see the reason for this, recall that the fixed costs of different activities are assumed to be the same for all firms, but the ability to cover these fixed costs depends on volume sold. High marginal cost firms will have small sales, and it is not profitable for them to incur the fixed costs of either exporting or FDI.

Firms with an intermediate level of unit costs will export but not undertake FDI, since the fixed costs of exporting are assumed to be less than those of FDI. Firms that have unit costs below some critical value switch from exporting to FDI, since they are able to sell at sufficient volume to cover the additional fixed costs. This critical value of unit costs is greater the higher are trade costs.

The model therefore predicts a size distribution of firms in each country, with firms engaging in a range of strategies some selling just in their domestic market, some exporting, and the largest engaging in FDI. The volume of affiliate sales relative to exports will be larger the higher are trade costs. Country differences can be added to the model, with results similar to those in the simpler model outlined above.

One further empirical prediction comes from this model, that the volume of FDI relative to exports will be larger the greater the degree of heterogeneity among firms. The intuition is that, other things being equal, exporting is a strategy associated with the middle of the distribution of unit costs, and increasing the dispersion of this distribution empties out this middle relative to the tails.

Empirical Evidence A number of papers have conducted empirical investigations of the hypotheses outlined above. An early and influential study by Brainard (1997) looks at the exports and affiliate sales of U.S.-based firms to 27 countries and in 63 industries. Trade costs are measured by freight and insurance charges and by import tariffs. Plant scale economies are proxied by plant-level employment, and firm-level scale economies by total nonproduction workers (capturing headquarters activities). Industry measures of R&D and advertising intensity are used to capture internalization advantages. Brainard finds that the share of affiliate sales in total foreign sales (affiliate sales plus exports) increases with trade barriers, transportation costs, and firm-level scale economies, and decreases with plant-scale economies, offering support to the predictions of the monopolistic competition model. Helpman, Melitz, and Yeaple (2004) undertake a similar exercise, with an extended data set and including a measure of the dispersion of firm size (and thereby dispersion of operating costs). Their results confirm those of Brainard, and additionally they find that dispersion has a negative effect on exports relative to affiliate sales, consistent with the predictions of their model of FDI with heterogeneous firms.

Policy Issues Many countries seek to attract FDI in the belief that it brings both employment and beneficial spillover effects, either through productivity spillovers or by stimulating related sectors. But at the same time, entry of FDI to a sector may crowd out local firms, and many countries restrict FDI inflow to sensitive sectors in order to protect local firms. How do these forces balance out?

Models of FDI and monopolistic competition are well placed to address these issues, because they allow for changes in the number of firms operating in a country and industry, and because they can be used to analyze linkages between related sectors (Rodriguez-Clare 1996; Markusen and Venables 1999). Suppose that a multinational firm enters a "downstream" industry (production of a final good). Its initial impact on local competitors is negative and, if the industry is monopolistically competitive, some of the firms in this industry may exit. However, the multinational will use inputs and these generate "backwards linkages," that is, create demand for intermediate goods supplied by upstream industries. Adjustment in these upstream industries may involve entry of new firms, and this can create a positive feedback mechanism. If these new upstream firms produce new (or better) varieties of intermediate goods, then costs in the downstream industry will fall (a "forward linkage"). This is a pecuniary externality the greater variety of intermediates benefits all firms in the downstream industry. It is then possible that entry of the multinational leads to a substantial increase in the industry's output essentially, it acts as a catalyst for formation of a cluster of activity. Whether this outcome happens depends on a number of parameters of the model and is more likely the more the multinational sources its inputs from the local economy rather than from imports, and the greater are the potential pecuniary externalities arising perhaps from variety effects, or also from direct knowledge spillovers or technological externalities.

These examples are illustrative of the rich set of possibilities that can be explored in models of FDI and monopolistic competition. There is potential for further work in several areas. For example, specification of sources of increasing returns needs to move beyond the simple characterization of fixed cost and constant unit cost. Richer modeling of firm heterogeneity would be desirable. A good deal of work needs to be done to further develop applications of the approach to policy, particularly in the context of understanding the contribution of FDI to the de-velopment of clusters of economic activity in developing countries and regions.

See also factor endowments and foreign direct investment; fixed costs and foreign direct investment; foreign direct investment and exit of local firms; foreign direct investment: the OLI framework; foreign direct investment under oligopoly; knowledge-capital model of the multinational enterprise; market size and foreign direct investment; monopolistic competition; outsourcing/offshoring; trade costs and foreign direct investment; vertical versus horizontal foreign direct investment

FURTHER READING

Barba Navaretti, Giorgio, Anthony J. Venables, Frank G. Barry, Karolina Ekholm, Anna M. Falzoni, Jan I. Haa land, Karen Helene Midelfart, and Alessandro Turrini. 2004. *Multinational Firms in the World Economy*. Princeton, NJ: Princeton University Press. Recent re view of theory and empirical literature on FDI and multinational firms.

Brainard, S. Lael. 1997. "An Empirical Assessment of the Proximity concentration Trade off between Multi national Sales and Trade." *American Economic Review* 87 (4): 520 44. Finds that sales by overseas affiliates of multinationals relative to exports increase with high trade costs, low investment barriers, plant level scale economies, and more similar countries.

Dixit, Avinash K., and Joseph E. Stiglitz. 1977. "Monop olistic Competition and Optimum Product Diversity." *American Economic Review* 67 (3): 297 308. The workhorse model of monopolistic competition.

Helpman, Elhanan, and Paul R. Krugman. 1985. *Market Structure and Foreign Trade*. Cambridge MA: MIT Press. A monograph on trade under imperfect compe tition.

Helpman, Elhanan, Marc J. Melitz, and Stephen R. Yeaple. 2004. "Export versus FDI with Heterogeneous Firms." *American Economic Review* 94 (1): 300 16. Employs the heterogeneous firm model to FDI and demonstrates the coexistence of firms that export and firms that under take FDI.

Markusen, James R., and Anthony J. Venables. 1999. "Foreign Direct Investment as a Catalyst for Industrial Development." *European Economic Review* 43 (2): 335 56. Model of the effect of FDI on local firms that

highlights how the profits of local firms can increase due to reduced input costs stemming from linkages to suppliers of intermediates.

. 2000. "The Theory of Endowment, Intraindustry and Multinational Trade." *Journal of International Economics* 52 (2): 209 34. Model demonstrating that multinationals are more likely to form when countries are more similar in terms of their relative (and absolute) endowments, as the data indicates is the case.

Rodriguez Clare, Andres. 1996. "Multinationals, Linkages, and Economic Development." *American Economic Review* 86 (4): 852 73. Model examining the conditions needed for linkage effects from FDI on the host country to be favorable.

ANTHONY J. VENABLES

■ foreign direct investment under oligopoly

An oligopoly is a concentrated industry in which only a few firms compete. The extreme case of oligopoly is duopoly, where just two firms exist in the market. The key qualitative characteristic of oligopoly is that there is significant interdependence between rival firms. Each firm is sufficiently large in relation to the overall product market for its actions (e.g., levels of research and development, plant investment, and pricing decisions) to have a significant effect on its rivals' profits.

Profit-maximizing firms in an oligopoly will recognize their interdependence and take account of the likely reactions of rival firms when making their own business choices. Accounting for strategic responses complicates the analysis of firm decision-making under oligopoly compared to monopoly and perfect or monopolistic competition, where firms typically act in isolation. Therefore, noncooperative game theory, which analyzes how independent, self-interested players behave in situations of mutual interdependence, is often used to predict the outcome ("equilibrium") of oligopolistic competition.

In an "international oligopoly," consumers are distributed over several countries, and a small number of firms populate the industry at a global level. Competition occurs on two levels: as well as producing output for each national product market, a key strategic decision for firms in an international oligopoly is the mode of market access. Firms choose whether to serve a given national product market through international trade (exports from a plant elsewhere) or local production. Foreign direct investment (FDI) is the name given to the funds that flow across national borders in order to establish production facilities abroad. A firm that undertakes FDI to set up plants outside its home country becomes a multinational enterprise (MNE), also known as a transnational corporation or a multinational corporation.

Empirical and Theoretical Points of Departure
Modern analysis of firms' plant location decisions in international oligopolies started with Stephen Hymer's famous PhD thesis, *The International Operations of National Firms* (submitted to MIT in 1960, but not published until 1976). Hymer began by noting that the predictions of the then-dominant theoretical approach to FDI flows, which treated FDI as an international flow of portfolio capital in a perfectly competitive global financial market, were strikingly contradicted by the data. In the traditional perspective Hymer critiqued, the rate of return to capital is given by its marginal product, which is itself a decreasing function of a nation's stock of capital relative to other factors of production. Therefore, the return to capital will tend to be low in capital-abundant (developed) countries and high in capital-scarce (developing) countries. Moreover, if capital is freely mobile internationally, then one would expect to observe FDI flowing from developed to developing countries.

Several representative features of the data (empirical "stylized facts") are strongly inconsistent with the traditional perspective. At the macro level, developed countries are the recipients of the majority of global FDI flows as well as being the majority source, of course. Developing nations are left largely out of the global FDI loop. In this vein, Hymer observed in 1960 that the most popular locations for postwar U.S. outward FDI were Canada and the United Kingdom, both developed nations whose capital stocks survived the war much less depleted

than those on the European continent. Moreover, two-way flows of FDI between pairs of countries ("cross-hauling") are common even at the industry level. Another robust and awkward empirical result comes from econometric studies of the macroeconomic determinants of FDI flows, which generally find little or no explanatory role for national measures of capital intensity or the return to capital. For excellent surveys of the empirical findings, see the introductory chapters of Barba Navaretti and Venables (2004), and Markusen (2002).

Hymer concluded that the traditional macroeconomic account of FDI flows under the assumption of perfect competition was inadequate, and that a microeconomic study of FDI emphasizing the special characteristics of MNEs and the markets they inhabit offered fresh hope. At the microeconomic level, Hymer and subsequent applied theorists highlighted two stylized facts for particular attention:

- MNEs are prevalent in relatively concentrated industries that seem to fit the theoretical category of oligopoly; and
- MNEs are most likely to be observed in industries where proprietary, knowledge-based assets (usually related to the production process or the product) are important. Moreover, within an industry, MNEs are generally the firms with the largest stocks of such assets. It seems intuitively obvious that industries where proprietary, knowledge-based assets are important should also be concentrated. The fixed costs of developing such assets generate significant economies of scale for incumbent firms, which means that there is only room for a few firms to operate profitably in the industry.

Moreover, because knowledge-based assets are nonrival within the firm (their so-called public good aspect), it is natural to expect multiplant firms to arise in industries where they are important. Of course, horizontal MNEs are a special case of multiplant firms, and Markusen (1984) examines their welfare consequences in a model where production requires both firm- and plant-level activities. The firm-level activity, which produces a joint input across all the firm's factories, might be research and development (R&D) investment. Markusen assumes a two-country world where each country hosts a single plant to produce the final good. The two plants may be owned by competing national firms or integrated within a single MNE. As Markusen shows, that organizational choice involves a trade-off between technical and allocative efficiency. Integrating the plants into a single MNE avoids duplication of firm-level activities and boosts profits, but it also creates a monopoly in the product market.

Dunning's popular OLI (ownership-location-internalization) framework suggests further reasons why FDI might be observed in knowledge-intensive industries. Given that a firm has chosen to produce abroad, Dunning argues that FDI is favored over licensing production to a foreign firm if the transactions costs of writing and enforcing contracts create sufficient "internalization advantages." The OLI approach also assumes that, in order to undertake FDI profitably, a firm must possess "ownership advantages" over its industry rivals (e.g., superior technology). We return to this point later.

Models of International Oligopoly Having established that knowledge-intensive industries are likely to be both oligopolistic and FDI-intensive, we now consider the direct relationship between FDI and oligopoly. The seminal early analysis of the implications of oligopoly for FDI behavior is Knickerbocker (1973). Knickerbocker documented follow-the-leader patterns in the data on firms' FDI behavior within given industries, most evident at intermediate levels of concentration. He claimed that tacit collusion lay behind this relationship: tacit collusion often leads to imitative, follow-the-leader behavior, and it naturally arises at intermediate levels of concentration ("loose-knit" oligopolies in his terminology). In contrast, if concentration is low, firms are too numerous to coordinate their activities; and if concentration is high (a "tight-knit" oligopoly), firms can coordinate to maximize industry profits and so surpass imitative behavior.

Graham (1998) uses game-theoretic techniques to formalize Knickerbocker's follow-the-leader story. In Graham's model, FDI is used as a threat to

maintain collusion. In an international duopoly where the firms originate from different countries (and, for simplicity, international trade is ruled out), each firm tacitly agrees not to enter its rival's home market via FDI, thereby granting its rival a domestic monopoly. If one firm breaks the understanding by undertaking FDI, its rival reciprocates, thus producing follow-the-leader behavior. (Graham's theoretical framework is an infinitely repeated Prisoner's Dilemma game.)

The modern game-theoretic literature has suggested other direct linkages between oligopoly and FDI. If technology diffuses slowly, then incumbent firms have first-mover advantages over potential entrants into the industry. These can facilitate preemptive FDI to keep host-country product markets concentrated and deter entry by local firms. The mechanism is that, compared to international trade, inward FDI allows foreign industry incumbents to compete more aggressively with local entrants on host-country product markets. The possibility of preemption creates a strategic motivation for FDI in addition to considerations of demand and cost. (Formal models are presented by Markusen 2002, chapter 4; Motta 1992.)

In this preemption story, FDI sustains market power, whereas the *threat* of FDI is key under tacit collusion. Both scenarios seem quite intricate, however: Do firms really think like that? The results of the formal models of tacit collusion and preemptive FDI are also highly sensitive to assumptions: for example, if the local firm in the preemption setup described earlier could move before the foreign industry incumbent, then it could enter the host-country market to preempt inward FDI! Such fragility is, however, a common feature of game-theoretic models.

Recent game-theoretic analyses of international oligopolies have uncovered further reasons for the emergence of MNEs. In an important paper, Horstmann and Markusen (1992) use a symmetric two-firm, two-country framework to examine firms' choices between horizontal FDI and exporting. In the model, the firms originate from different countries, and competition takes place in two stages. First,

the firms choose between not entering the industry and entering either as a national firm or as a horizontal MNE. If it enters, a firm incurs a firm-specific fixed cost, which represents its headquarters and other joint-input activities such as R&D. Firms also incur a plant-specific fixed cost on each factory they establish. Therefore, choosing between entering as an MNE or a national firm (i.e., FDI versus exporting to the foreign market) involves a trade-off between fixed and marginal costs. FDI entails the fixed cost of an additional plant, but it allows the firm to "jump" the international trade cost that must be paid on exports. After having chosen the location of production, the firms compete (in Cournot fashion) on both national product markets. As is conventional, Horstmann and Markusen solve their model backwards to isolate its subgame perfect Nash equilibrium, thus capturing the idea that firms consider how their FDI choices will affect subsequent product-market competition.

Two results from the Horstmann-Markusen model are especially noteworthy. First, MNEs become more likely to arise in equilibrium as the firm-specific fixed cost rises relative to that for plants. This result echoes Markusen (1984): integrating national firms into an MNE economizes on firm-specific costs. Moreover, FDI cross-hauling occurs in equilibrium when the national markets are large enough to support the total fixed costs of two MNEs. This explanation of cross-hauling reciprocal FDI flows within a given industry is a major contribution because such flows are both empirically common and difficult to rationalize within other frameworks, such as OLI. Recall that a precondition for FDI in the OLI framework is that a firm possess "advantages" over its rivals. Therefore, FDI cross-hauling within an international duopoly suggests that firm A has advantages over B and B over A clearly an impossibility.

A second key result from Horstmann and Markusen (1992) is their prediction on how changes in international trade costs affect firms' choices between horizontal FDI and exporting. Because the benefit of FDI to firms is the avoidance ("jumping") of trade costs, Horstmann and Markusen find that national firms tend to displace MNEs as trade costs fall. This

prediction seems strikingly counterfactual, as we discuss later.

Markusen and Venables (1998) extend the analysis of firms' FDI decisions by allowing for entry into the industry and differences between the two host countries. They find that MNEs tend to displace national (exporting) firms in equilibrium as countries become more similar (both in size and in relative factor endowments), and as world income grows. These results fit the stylized facts well because a large proportion of global FDI flows between "similar" developed countries. Intuitively, similarities between the host countries encourage two-way FDI flows because they mean that MNEs from different countries are able to compete on similar terms.

A final area where oligopoly theory has been applied to FDI is in the analysis of horizontal cross-border mergers and acquisitions (M&As). Quantitatively, international M&As are more important in FDI flows than greenfield investment, the name given to the building of new production facilities abroad. Moreover, analysis of international M&As helps address a well-known paradox. As confirmed by Horstmann and Markusen (1992), intuition suggests that reductions in international trade costs should discourage FDI in favor of exports and, within FDI, the share of vertical should grow because it involves some international trade, whereas horizontal FDI involves none. There has been dramatic growth in global FDI flows in recent decades, however, and the majority of FDI remains horizontal, despite the large reduction in international trade costs, due in part to ongoing trade liberalization.

Neary (2006) considers possible resolutions of this paradox. Unlike horizontal greenfield FDI, he shows that trade liberalization can increase the profitability of horizontal mergers. The profitability gain from a merger arises because it removes the competition that would otherwise have occurred between the participating firms. Moreover, competition in the global market is more intense, the lower are trade barriers. With high trade barriers, import competition is limited and, for example, isolated national firms can earn almost monopoly profits at home. Therefore, the profitability gain from reducing competition through merger/acquisition rises as trade costs fall, and we expect trade liberalization to encourage horizontal cross-border M&As.

Another potential resolution of the paradox is to allow for intrabloc trade liberalization, for example, the launch of the European Single Market or the North American Free Trade Agreement (NAFTA). These liberalizations might make horizontal inward FDI from the rest of the world to serve the whole bloc profitable, where previously inward FDI to serve primarily just one bloc member was not.

This survey has aimed to show how oligopoly theory can cast light on strategic motivations for FDI that are absent in other market structures. In the process, some empirical observations such as FDI cross-hauling at the industry level become easier to explain. Finally, we should note that although this survey has ignored policy issues, there are lively debates concerning oligopoly and FDI. One, for example, concerns whether large MNEs are able to "divide and rule" national governments and cause a "race to the bottom" in corporate taxes. Such controversies highlight what a fruitful area this is for further research.

See also factor endowments and foreign direct investment; fixed costs and foreign direct investment; foreign direct investment and exit of local firms; foreign direct investment: the OLI framework; foreign direct investment under monopolistic competition; knowledge-capital model of the multinational enterprise; market size and foreign direct investment; outsourcing/offshoring; trade costs and foreign direct investment; vertical versus horizontal foreign direct investment

FURTHER READING

Barba Navaretti, Giorgio, and Anthony J. Venables. 2004. *Multinational Firms in the World Economy*. Princeton, NJ: Princeton University Press. A recent and important monograph on multinational enterprises.

Dunning, John H. 1977. "Trade, Location of Economic Activity, and the MNE: A Search for an Eclectic Ap proach." In *The International Allocation of Economic Activity: Proceedings of a Nobel Symposium held at Stockholm*, edited by B. Ohlin, P. O. Hesselborn, and

P. M. Wijkman, eds. London: Macmillan, 395 418. Definitive statement of the OLI framework.

Graham, Edward M. 1998. "Market Structure and the Multinational Enterprise: A Game Theoretic Ap proach." *Journal of International Business Studies* 29 (1): 67 83. Very readable application of game theory to study of MNEs.

Horstmann, Ignatius, and James R. Markusen. 1992. "Endogenous Market Structures in International Trade (Natura Facit Saltum)." *Journal of International Eco nomics* 32 (1 2): 109 29. Rigorous analysis of the FDI/ exporting choice in an international duopoly.

Hymer, Stephen H. 1976. *The International Operations of National Firms: A Study of Direct Foreign Investment.* Cambridge, MA: MIT Press. The starting point of the modern theory of FDI.

Knickerbocker, Frederick T. 1973. *Oligopolistic Reaction and Multinational Enterprise.* Boston, MA: Harvard University Press. An important early work on FDI under oligopoly.

Markusen, James R. 1984. "Multinationals, Multiplant Economies, and the Gains from Trade." *Journal of In ternational Economics* 16 (3 4): 205 26. A key early work on horizontal FDI.

——. 2002. *Multinational Firms and the Theory of In ternational Trade.* Cambridge, MA: MIT Press. Survey of the author's contributions to the theory of FDI under oligopoly.

Markusen, James R., and Anthony J. Venables. 1998. "Multinational Firms and the New Trade Theory." *Journal of International Economics* 46 (2): 183 203. Extends Horstmann and Markusen (1992) to allow for country differences and entry into the industry.

Motta, Massimo. 1992. "Multinational Firms and the Tariff Jumping Argument: A Game Theoretic Analysis with some Unconventional Conclusions." *European Economic Review* 36 (8): 1557 71. A formal model of pre emptive FDI.

Neary, J. Peter. 2006. "Trade Costs and Foreign Direct Investment." CEPR Discussion Paper No. 5933. Lon don: Center for Economic Policy Research. Addresses the empirical paradox of growing horizontal FDI flows in the presence of falling trade costs.

BEN FERRETT

■ foreign equity restrictions

A foreign equity restriction is a measure that imposes more stringent requirements on foreign investors than domestic investors and thereby influences de cisions about where to invest and in what form. In some economies, such as Switzerland, private com panies impose limits on foreign holdings of their own shares (Bailey, Chung, and Kang 1999), but in most cases foreign equity restrictions are the result of government policy.

The aim of regulatory restrictions on foreign eq uity ownership is to maintain the independence of domestic firms for reasons of sovereignty, national security, and economic stability. In particular, the measures are often intended to prevent the acquisi tion of controlling stakes in domestic companies by large multinational corporations, over which na tional governments and other stakeholders have little influence. Sectors that are frequently subject to for eign equity restrictions are telecommunications and media, air and maritime transportation, finance, public utilities, and defense electronics.

Resource Allocation Implications Foreign direct investment (FDI) is a form of international eco nomic integration that can bring substantial gains to both investor and investment-receiver by fostering intrafirm trade and transactions in intangible assets, such as knowledge and product brands. Beyond the standard gains from trade, FDI inflows can provide dynamic spillovers from technology transfer and skill-building. These benefits are especially impor tant in developing countries where advanced tech nology and managerial expertise are in short supply.

Barriers to FDI reduce or eliminate these benefits by distorting the allocation of capital among different economies, between foreign and domestic invest ment, among different sectors, and between tem porarily held portfolio and longer-term direct in vestment. As a result, products and services may cost more than they would without the foreign equity restrictions and assets may not be used in the most productive way. The effects may manifest themselves through a variety of channels, such as higher prices, less consumer choice, lower capital stock, and lower productivity. The resulting costs of the foreign equity

restrictions in terms of forgone economic welfare improvements have to be weighed against their noneconomic benefits in terms of sovereignty, security, and stability to arrive at an overall assessment of the rationale for such measures.

Types of Restrictions and Measurement Measures to control or influence FDI range from simple requirements that investments be registered and screened, with virtually automatic approval, to complete bans on foreign ownership in some sectors or countries. Restrictions also vary with respect to the level at which they are applied. For example, measures may affect market entry, ownership and control, or operations. Another classification of re-

strictions pertains to the way in which they are applied for example, through legislation that clearly specifies ownership limits, or case-by-case assessments of whether entry will be allowed and the conditions that may apply (Hardin and Holmes 1997).

Assessing the relative prevalence and restrictiveness of different foreign equity restrictions presents major challenges, similar to those encountered when trying to measure other nontariff barriers. Relatively simple ways of summarizing the extent of foreign equity restrictions are based on counts of the number of measures in a sector or economy, or on indicators that reflect the proportion of investments covered by

Table 1

Coefficients on FDI restrictions (maximum 1.0)

Type of restriction	Scores
Foreign equity limits	
No foreign equity allowed	1
1 to 19% foreign equity allowed	0.6
20 34% foreign equity allowed	0.4
35 49% foreign equity allowed	0.3
50 74% foreign equity allowed	0.2
75 99% foreign equity allowed	0.1
no restriction but unbound	0.05
Screening and approval	
Investor must show economic benefits	0.2
Approval unless contrary to national interest	0.1
Notification (pre or post)	0.05
Other restrictions	
Board of directors/managers	
majority must be nationals or residents	0.1
at least one must be national or resident	0.05
must be locally licensed	0.025
Movement of people	
no entry	0.1
less than one year	0.075
one to two years	0.05
three to four years	0.025
Input and operational restrictions	
domestic content must be more than 50%	0.1
other	0.05
Total[a]	Between 0 and 1

Source: OECD, adapted from Hardin and Holmes (1997).

[a]If foreign equity is banned, then the other criteria become irrelevant, so that the index is at 1.0. It is possible that various scores sum to slightly more than 1.0 when foreign equity is not totally banned, and in such cases, the index is capped at 1.0.

barriers (Hoekman 1995). Although such frequency and coverage indexes have the advantages of simplicity and lack of arbitrariness, some measures are more constraining than others. It therefore seems desirable for analysts to attempt to weight different restrictions according to their significance, even though such a procedure entails some subjective judgments and possible errors (Golub 2003).

One weighting scheme that has been used by Organisation for Economic Co-operation and Development (OECD) researchers is shown in table 1. The overall restrictiveness of FDI barriers is derived for each industry and country, based on regulations in three areas: equity limits, screening and approval, and other restrictions. Area scores are added up to yield an overall indicator that ranges from 0 to 1. Equity limitations and screening are related to an investor's right of establishment, whereas measures affecting management structures, movement of people, and operational restrictions pertain to national treatment of established firms.

This line of analysis focuses on statutory restrictions, but many countries retain informal public or private barriers to foreign equity acquisition, which can in some cases be more restrictive than formal regulations. Nonformalized FDI impediments are by nature difficult to capture and compare, however.

Policy Reform Many economies have liberalized their foreign investment policies since the 1980s such that equity restrictions in manufacturing have virtually disappeared in OECD countries. Developing and emerging economies maintain on average more restrictive policies, but large variations across countries and sectors exist. While service sector privatizations have led to increased market openness in many economies, reforms have often involved limits on acquisitions by foreigners. Yet there are some notable differences between OECD and non-OECD countries. In particular, transportation and tourism are relatively less restricted in non-OECD countries, while electricity, wholesale and retail distribution, and finance are relatively highly restricted in non-OECD countries (Koyama and Golub 2006).

The general, broad-based trend toward removal of foreign equity restrictions suggests that policymakers have increasingly been persuaded of the merits of inward FDI over concerns related to sovereignty, security, and stability, as well as the partisan interests of entrenched lobbies. From an economic point of view, nondiscrimination between domestic and foreign investors is generally the best policy in the absence of a clear-cut market failure or threat to national interests. Neutrality involves both right of establishment for foreign firms and national treatment of such firms once they are established. Right of establishment signifies that there are no discriminatory obstacles to foreign greenfield investment or mergers and acquisitions, while national treatment involves nondiscrimination in conducting business (Golub 2003).

Progress toward liberalization of remaining restrictions can be facilitated by continued discussion in international forums on development of codes, guidelines, and best practices in areas with direct impact on public- and private-sector governance practices. Moreover, reforms do not necessarily have to involve the immediate and complete removal of foreign equity restrictions. In some cases, more direct measures that, for example, clearly specify quantity or value constraints on foreign investors' market access or operations will tend to achieve the policy goal at a lower cost than a regulation that tries to achieve the same objective but in a less direct and transparent way.

See also agglomeration and foreign direct investment; foreign direct investment (FDI); foreign market entry; international investment agreements; joint ventures; mergers and acquisitions; multinational enterprises; trade-related investment measures (TRIMs)

FURTHER READING

Bailey, Warren, Y. Peter Chung, and Jun koo Kang. 1999. "Foreign Ownership Restrictions and Equity Price Premiums: What Drives the Demand for Cross Border Investments?" *Journal of Financial and Quantitative Analysis* 34 (4): 489 511. Assesses the effects of foreign equity restrictions imposed by private companies on equity prices.

Golub, Stephen S. 2003. "Measures of Restrictions on In ward Foreign Direct Investment for OECD Countries." *OECD Economic Studies* 36: 85 115. Provides a dis cussion of the measurement of FDI restrictions and derives restrictiveness indicators for OECD countries.

Hardin, Alexis, and Leanne Holmes. 1997. *Service Trade and Foreign Direct Investment.* Canberra: Australian Productivity Commission. Analyzes a broad range of issues with respect to FDI restrictions in services sectors.

Hoekman, Bernard. 1995. "Assessing the General Agree ment on Trade in Services." In *The Uruguay Round and the Developing Countries*, edited by Will Martin and L. Alan Winters. Washington DC: World Bank, 327 64. Analyzes the GATS partly based on frequency counts of foreign equity restrictions.

Koyama, Takeshi, and Stephen S. Golub. 2006. "OECD's FDI Regulatory Restrictiveness Index: Revision and Extension to More Economies." OECD Economics Department Working Papers No. 525. Paris: OECD. Provides estimates of FDI restrictiveness for more than 40 OECD and emerging economies.

PETER WALKENHORST

■ foreign exchange intervention

Foreign exchange intervention refers to the purchase and sale of currencies in the foreign exchange market by a country's monetary authority, such as its central bank. In many countries with managed exchange rates, intervention is used routinely to target the for eign exchange value of the country's currency. For example, a central bank with a currency whose value is dipping below its target range typically will sell some of its holdings of foreign currency in exchange for purchases of its own currency. This process reduces the supply of the country's own currency and thereby increases its value in the foreign exchange market.

In countries without exchange rate targets, foreign exchange intervention is used at times to try to in fluence other aspects of the foreign exchange market or to directly alter domestic macroeconomic condi tions. Many central banks intervene in an attempt to dampen the volatility of the exchange rate, to slow the rate of change of the currency's foreign exchange value, to calm disorderly markets, or to manage the official holdings of foreign exchange reserves. In a few countries, foreign exchange intervention is a standard tool for addressing important domestic macroeco nomic concerns, such as the rate of inflation. Some of these domestic concerns may be linked only indi rectly to the foreign exchange market itself. In prin ciple, a monetary authority can use intervention to achieve any number of objectives.

When managing foreign reserves or influencing any aspect of the foreign exchange market is the ob jective of intervention, the monetary authorities face several choices in its implementation. Most impor tant, a monetary authority must decide to what extent the foreign exchange intervention should be allowed to affect domestic monetary conditions. Foreign ex change intervention can be undertaken in tandem with offsetting domestic monetary operations. An intervention's effect on domestic monetary conditions thereby will be "sterilized"   that is, there will be no net direct effect on domestic monetary conditions. Alternatively, intervention can be undertaken by itself, without countervailing domestic operations. Such intervention is called "unsterilized" intervention.

A monetary authority must also choose how much information about its intervention to divulge to foreign exchange market participants. It may in tervene in secret, it may reveal some of its informa tion about intervention, or it may intervene openly. Finally, a monetary authority may choose to inter vene in concert with others, in which case its inter vention is said to be "concerted," or it may intervene unilaterally. The efficacy of foreign exchange inter vention is thought to vary depending on these key aspects of its implementation   whether it is steril ized or unsterilized, secret or transparent, and uni lateral or concerted.

Effectiveness of Foreign Exchange Interven tion Intervention is quite effective in modifying the foreign exchange value of currencies when it is un sterilized. Since it is not offset by countervailing domestic operations, unsterilized intervention alters the amount of reserves in the banking system. When a central bank, for example, sells some of its foreign currency holdings, the reserves in the banking system

decline correspondingly. Foreign exchange sales themselves are thus a form of monetary policy tightening. This type of intervention has a direct effect on the foreign exchange value of a country's currency: tighter monetary conditions strengthen a currency's foreign exchange value. Its effectiveness is apparent even for currencies with high turnover in deep markets. When exchange rate concerns dominate monetary policy, foreign exchange intervention is effective.

Sterilizing foreign exchange intervention means preventing it from bringing about a domestic monetary condition change the very thing that would have the most direct effect on the currency's value. Therefore, in order for sterilized intervention to have an effect on the exchange rate, its effectiveness must come about through some other channel.

One important potential channel is known as the signaling channel. In this channel, intervention affects exchange rates when it signals a change in future monetary policy. In theory, such signals are credible (and intervention effective) only if intervention tends to be followed by consistent changes in monetary policy. Thus the signaling channel allows for an effect on exchange rates when the intervention, though initially sterilized, is expected ultimately to be unsterilized.

The microstructural approach to foreign exchange has recently provided other new insights into the efficacy of intervention. Microstructure models focus on the informational structure of foreign exchange markets. Because they are explicit about the informational structure, these models are well-equipped to show how an individual agent (such as the monetary authority) reveals its information when it initiates a transaction. This makes the microstructural approach particularly suitable both for understanding the role that intervention might play in coordinating exchange rate expectations and for sorting out some of the key aspects of the intervention's actual implementation. For example, microstructure models can provide precise answers to questions such as why intervention is carried out secretly, or why intervention, rather than an announcement, is used to convey information. The

intervention's optimal implementation is shown in these models to depend on the precise nature of the informational asymmetries in the market.

Practice of Foreign Exchange Intervention In most mature economies with floating exchange rates and open capital markets, intervention now occurs only infrequently. Although central banks engaged in more frequent unilateral and concerted intervention during some earlier periods, such as the 1980s, their intervention in the early 21st century was quite limited. For the central banks that issue the major international currencies, intervention when it does occur is usually small relative to the overall size of the foreign exchange market. It also typically takes the form of sterilized intervention. The only notable exception is the Bank of Japan, which has intervened frequently, and sometimes extensively. The European Central Bank and the U.S. Federal Reserve Board, for example, may let years go by without carrying out any intervention at all; and, when they do intervene, their intervention is sterilized routinely. Even the Bank of Canada, which had a policy of regular and automatic intervention until the mid-1990s, countenances intervention only in exceptional circumstances. For mature economies, extended central bank absence from the foreign exchange market is the norm.

The monetary authorities in many emerging economies, on the other hand, still intervene heavily in the foreign exchange market. In 2000 2006, the monetary authorities in China, in several of the oil exporting countries, and in a handful of emerging market countries sold off large quantities of their own currencies. In doing so, they accumulated immense holdings of foreign currency reserves.

Intervention among emerging market countries is most prevalent in those with exchange rate targets. Their intervention in some cases is unsterilized, and the corresponding changes in monetary conditions affect the exchange rate directly. In other cases, particularly where intervention occurs in a setting with significant barriers to capital mobility, it is sterilized. This combination of sterilized intervention and barriers to capital mobility is thought to enable a country to maintain a relatively weak exchange rate

without engendering domestic inflation. It is also exactly this combination of sterilized intervention and capital controls that so frustrates the trading partners of a few emerging market countries and provokes accusations of excessive currency manipulation.

Research Challenges The practice of foreign exchange intervention has moved along in fits and starts. It has varied across both time and countries in nearly every way in its objectives, in its implementation, and in its perceived effectiveness. This variation is mirrored in the dissonance of research on intervention, which as of the first decade of the 21st century provided little in the way of sure-footed guidance to monetary authorities. (In fact, the extent of any consensus arguably diminished since earlier decades, when a deeper skepticism about intervention prevailed.) Researchers face daunting empirical and theoretical challenges, and definitive conclusions remain a long way off.

On the empirical front, researchers face a scarcity of reliable, detailed data on actual intervention for all but a few countries. Foreign exchange market news reports have been used in lieu of actual intervention, but these reports have been found to be inaccurate. They can be used to measure at best only what the exchange rate response is to reports of intervention, not actual intervention. The reports are particularly unhelpful in exploring the role of secrecy in intervention. In addition, key pieces of intervention's implementation, such as whether or not it is sterilized or is concerted, often are not directly observable. This is problematic since theory tells us that intervention's effectiveness depends on these key pieces of its implementation. Moreover, intervention policy even its objective does not appear to be stable over time. This structural instability is generally, though not always, difficult to detect. When undetected, it means that atheoretical empirical studies are likely to conflate measures of effectiveness from different policy eras. Such conflated measures are difficult to interpret.

Empirical work on the intervention's exchange rate effectiveness also faces a serious simultaneity problem: foreign exchange intervention is carried out in response to changes in the exchange rate and in response to the conditions that affect the exchange rate. Thus intervention's effect on the exchange rate must be disentangled from the exchange rate's effect on intervention. Credible empirical studies must begin with plausible strategies to identify the effect of intervention by itself and that is a tall order. The most common identification strategy has relied on assumptions about the timing of the effects: the strategy is to assume that intervention affects the exchange rate instantly, while the exchange rate affects intervention only slowly. Unfortunately, to the extent that monetary authorities "lean against the wind," their response to the exchange rate is automatic. Thus the assumption of a slow response is often rendered implausible and the problem of disentangling the effects reemerges.

Further complicating the empirical work are the important theoretical challenges to a clear understanding of foreign exchange intervention. For example, while some models provide good descriptions of the implications of asymmetric information, only a few address the source of the informational asymmetry. It is easy enough, at one level, to say that a monetary authority has privileged information about its own intervention objectives and about future monetary policy. However, that is a remarkably static explanation for work in a field that generally demands attention to dynamics. It calls for further explanation of why such an informational advantage might not erode over time. A dynamically consistent account of the informational asymmetry will pave the way for a more complete understanding of the role of foreign exchange intervention.

See also asymmetric information; capital controls; capital mobility; equilibrium exchange rate; exchange rate regimes; exchange rate volatility; fear of floating; impossible trinity; international reserves; money supply; new open economy macroeconomics; reserve currency; sterilization; vehicle currency

FURTHER READING

Dominguez, Kathryn M. E. 2003. "The Market Micro structure of Central Bank Intervention." *Journal of International Economics* 59: 25 45. Explores the intraday

timing of intervention's effectiveness and the role of news reports of intervention.

Edison, Hali J. 1993. "Effectiveness of Central Bank In tervention: A Survey of the Literature after 1982." *Spe cial Papers in International Economics* 18: 1 68. Inter national Financial Section, Princeton University. A benchmark review of the early literature on central bank intervention.

Popper, Helen, and John Montgomery. 2001. "Informa tion Sharing and Central Bank Intervention." *Journal of International Economics* 55: 295 316. Models a novel source of the central bank's privileged information and is one of the few papers to examine explicitly some of in tervention's welfare effects.

Sarno, L., and M. P. Taylor. 2001. "Official Intervention in the Foreign Exchange Market: Is It Effective and, If So, How Does it Work?" *Journal of Economic Literature* 39: 839 68. Updates Edison (1993), also explores inter vention's role in addressing a coordination failure.

Vitale, Paolo. 1999. "Sterilized Central Bank Intervention in the Foreign Exchange Market." *Journal of Interna tional Economics* 49: 245 67. A classic microstructural approach to intervention's signaling role. Addresses the issue of secrecy when the central bank's target differs from the fundamentals.

HELEN POPPER

■ foreign market entry

Foreign market entry is an important strategy for multinational enterprises (MNEs) to expand their global reach and establish an international network of interdependent business units. Within this net work, each subsidiary has a specific role, such as to provide access to local resources or markets. The design of an entry strategy has to match the needs and resources of the MNE with the opportunities and constraints in the local environment. Entry strategies are thus adapted to the local resource endowment, market demand, and institutional environment, though the degree of such adaptation varies across firms, industries, and locations.

Setting up a new business operation requires several strategic decisions, including the entry mode, location, timing, marketing, human resources, and logistics. Most scholarly research has focused on entry modes: export, contractual cooperation, or foreign direct investment (FDI). The FDI mode is further distinguished by the investors' share of equity ownership, and whether to build an operation from scratch, a greenfield project, or acquire an existing firm. These decisions are interdependent though economic analysts often prefer to analyze them ce teris paribus to generate more parsimonious theo retical models. Key parameters influencing the entry strategy vary with the objectives of the project; they are here illustrated primarily with a focus on market-seeking FDI.

Foreign entries are undertaken for a reason that is normally specified before the entry strategy is de signed. The motives of FDI relate to the develop ment and exploitation of resources of the MNE across its operations. Hence any specific entry deci sion cannot be viewed in isolation but has to be considered in relation to the overall strategic posture of the MNE. The MNE's global strategy thus sets the framework for specific foreign entries. This article considers the following aspects of foreign entry: entry mode, location, timing, marketing, human resource management, and logistics.

Entry Modes Foreign entry modes are classified in multiple ways. First, nonequity modes are dis tinguished from equity modes; the former include exports with and without trade intermediaries, as well as contractual forms of business such as licens ing, franchising, outsourcing, and turn-key projects. Equity modes are distinguished by the ownership stake taken by the foreign investor: portfolio invest ment, joint venture (JV), and wholly owned sub sidiary (Young, Hamill, Wheeler, and Davies, 1989; Welch, Benito, and Petersen 2007).

Portfolio investment involves the acquisition of a small equity stake as a financial investor, yet without influencing the strategy of the firm. Joint ventures are created by two or more partner firms each contributing resources and sharing the con trol of the operation. They are differentiated by ownership levels taken by the foreign investor; the higher the equity stake and the more dispersed the

residual ownership the more the investor has control over the operation. Wholly owned subsidiaries provide the investor with full control.

Theoretical models of the choice of foreign equity focus on the internalization of transactions and control over operations (Buckley and Casson 1996; Brouthers and Hennart 2007). Transaction costs economics and internalization theory suggest that a JV would be chosen if three conditions are met: (1) the new business unit depends on resource contributions from two or more firms, (2) the transfer of these resources or the expected benefits for the investors are subject to high transaction costs, and (3) it is not feasible for the entire parent firms to be integrated into one firm, for instance because they are big relative to the envisaged project, or one of them is a state-owned enterprise.

High transaction costs are in particular likely to arise for investors who transfer intangible assets. This issue is particularly pertinent for MNEs because their ability to achieve competitive advantages in foreign markets is often based on the deployment of intangible assets. They face asymmetric information regarding the content, value, and usage of these assets, which are classic sources of market failures. Thus the more markets are characterized by information asymmetries, the more likely MNEs prefer to internalize an operation and establish a wholly owned subsidiary. Moreover, Kogut and Zander (1993), and other scholars, argue that knowledge to be transferred is often tacit and thus requires "learning by doing" and therefore an intraorganizational mode of transfer. This requirement provides an alternative explanation for why knowledge-intensive firms would abstain from licensing or JVs as modes of foreign entry. Thus the more MNEs compete on the basis of technology, brand names, or other intangible assets, the more likely they would enter foreign markets with wholly owned subsidiaries.

A different source of transaction costs is interdependence of business activities and thus high asset specificity. These transaction costs lead in particular to vertical JVs in industries that require large investments in assets specific to a business relationship, as in the mining and processing of certain natural resources.

Yet, transaction costs also depend on the institutional environment governing the market, notably the feasibility and cost of enforcing contracts. Where these institutions are weak, investors would abstain from relying on contracts; at the same time, local partners may be helpful in interacting with other local businesses in such contexts. Joint ventures may provide an avenue to operate in unfamiliar contexts, especially where market supporting institutions are weak.

A different distinction of investment modes concerns the make-or-buy decision, namely the establishment of a greenfield operation versus an acquisition of (a stake in) an existing company. This decision is primarily driven by the investor's need for local resources: an acquisition provides local organizationally embedded resources, such as human capital and networks with local authorities. A greenfield operation, in contrast, allows investors to create a new operation from scratch according to their own designs, and thus to replicate organizational procedures and practices. The greenfield option is preferred in particular by investors whose competitive advantages are grounded in the firm's organizational structure and culture. Acquisitions are preferred by those who need complementary local resources.

Location Location concerns both the choice of country to invest in and the selection of a specific site. For resource- or efficiency-seeking projects aiming to create an export platform, the primary concern is costs of production. Hence the key determinants for location decisions are the costs and quality of the local factor endowment. Entrants would consider the specific inputs that they require for their operations, including natural assets such as cost of the local workforce and natural resources and "created assets" such as intermediate goods, human capital, and infrastructure.

Created assets are often of primary concern in industrialized countries, but they are also of increasing importance in emerging economies. Created assets are often provided by other businesses, which is one cause of the agglomeration of FDI:

foreign entrants invest where a strong community of local and foreign-owned businesses already exists.

Market-seeking investors are primarily concerned with access to distribution channels and potential customers. They identify the relevant market and then seek a central location for sales, marketing, and distribution operations. In some industries, the actual production needs to be located close to the customer or the point of consumption, notably in service industries such as hotels and financial services, and for manufactured goods that face high transportation costs. Thus location decisions by market-seeking investors are a function of the expected market size and of transportation costs of bringing products to the customer. For consumer goods, market demand may be predicted based on demographic data such as per capita incomes and population size. For business-to-business operators, size and growth of the local customer industry are crucial. A special case is "follow-the-customer" entrants, which set up operations to best supply a specific customer, as commonly observed, for instance, in the automotive industry.

In addition to the primary considerations of factor costs and markets, location decisions are often moderated by the institutional environment. For instance, investors are concerned about regulation of the industry, political risk, and law enforcement, especially with regard to intellectual property rights. Moreover, they may prefer locations in cultural proximity because cultural differences between existing and new operations may reduce the efficiency of communication and coordination.

Location decisions may be made in two stages, first the country and then the site, but businesses may also compare alternative sites in different countries. The relevant parameters for both between-country and within-country location choice tend to be similar. MNEs may shortlist alternative sites and then negotiate with the pertinent national and local authorities in parallel. At this stage, investment incentives such as provision of specific local infrastructure or subsidies and financial incentives may tip the balance for a particular site, especially if competition for FDI is strong.

Timing of Entry Market-seeking foreign investors normally aim to be number one or two in their industry or market segment, especially if the industry tends toward oligopolistic market structures. Such market leadership may best be achieved by taking an early lead, or by taking over a local market leader (Lieberman and Montgomery 1988). Thus market-oriented investors often pursue first-mover advantages, while others enter the host country shortly after the first mover, aiming to challenge the first mover before its position is too strong.

Early entrants may attain a lead in building reputation and consumer loyalty, and in establishing relationships with major suppliers and customers. First movers may thus be able to lock their partners and customers into a relationship, and raise barriers to entry for later entrants. Moreover, first movers may build goodwill with local authorities, slide down the learning curve, and acquire unique local resources, such as distribution channels, local brands, and raw material sources.

Followers, on the other hand, may benefit from a less uncertain business environment, from observing the experiences of the first mover, and from customers and local authorities already familiar with the product and the practices of the industry. In particular, "fast seconds" may benefit from these advantages, while challenging the first mover before the market structure has stabilized. Empirical evidence suggests that first movers can maintain their leadership position if they continuously commit resources and focus on learning about the local environment. Yet evidence also points to many first movers who did not succeed in creating sustained market leadership.

Related to timing is the question of how fast to accelerate commitment to a market. In some industries, heavy up-front investment is required to establish a strong position in view of (potential) competitors, as for branded consumer goods, or because of the capital intensity of the industry, as for oil exploration. In other industries, a "platform strategy" may be appropriate as investors establish a small foothold from which to observe the local industry and to flexibly react to business opportunities if and when they emerge.

Marketing The marketing literature on foreign entry has focused on standardization versus local adaptation (Meyer and Tran 2006). A standardization strategy would use the same product, brand, and processes in all countries. Advantages of such a strategy include economies of scale in product development, production, and marketing. It is most likely to be appropriate in industries that are technology intensive and face little variation in consumer preferences. In contrast, a local adaptation strategy would emphasize localization of products to accommodate local needs and preferences. This may in particular include the creation or acquisition of local brands for the mass market. Such localization is important for instance in the food and beverage industry.

Standardization and localization are not exclusive, as MNEs may combine a high degree of standardization for their internal processes, such as product development and production, with adaptation of products and customer interfaces, notably marketing practices. Others pursue a multi-tier strategy and offer both an international product positioned in the premium segment and a local brand aimed at the mass market. Such a strategy allows synergies, for instance, in the use of distribution channels.

Human Resources and Logistics A foreign entry depends on qualified and motivated people to implement the strategy. Human resources management is particularly important to manage knowledge sharing within the organization, both to transmit key organizational practices to the new operation and to inform decision-makers at headquarters about the local business. Thus along with the establishment of a new subsidiary, expatriate managers have to be selected and prepared for their assignment, and local staff needs to be recruited and trained. Expatriates play a pivotal role in this process, as do training programs for local staff that may include learning by doing in other operations of the MNE. Other human resource management challenges include leadership of a workforce in a culturally different context and adaptation of systems for recruitment, performance assessment, and remuneration of local staff.

An important aspect of a foreign entry strategy that is little analyzed in the literature but of great importance for management practice is logistics. Lower labor costs are valuable for a business only if the products can be transported to the customer in good time at acceptable costs. Modern transportation infrastructure and information technology systems are designed to allow MNEs to optimize the integration of their internal operations as well as supplier relations. Specialist intermediaries offer services that may include not only warehousing, shipping, and door-to-door delivery but also processing of customer orders and identification of appropriate suppliers. A foreign entry often triggers changes in these systems, especially if it involves the relocation of production.

Interdependencies An entry strategy consists of many elements; entry strategy is, therefore, a multidimensional construct. Any decision has to take account of other dimensions. Take, for example, the choice of entry mode. If timing and speed of entry are crucial for an investor   as they are for those pursuing a first-mover advantage   an acquisition or a JV may offer quick market access. However, an acquisition poses greater challenges for the integration of the local operation with the global company in terms of, for instance, marketing, logistics, and human resource management. Thus foreign market entry decisions have to reflect the complex interdependence of multiple dimensions.

See also foreign direct investment: the OLI framework

FURTHER READING

Brouthers, K. D., and J. F. Hennart. 2007. "Boundaries of the Firm: Insights from International Entry Mode Re search." *Journal of Management* 33 (3): 395 425. A critical review of the literature on ownership modes.

Buckley P. J., and M. C. Casson. 1996. "An Economic Model of International Joint Venture Strategy." *Journal of International Business Studies* 27: 849 76. A formal analysis of operation modes for different types of operations.

Kogut, B., and U. Zander. 1993. "Knowledge of the Firm and the Evolutionary Theory of the Multinational Corporation." *Journal of International Business Studies*

24 (4): 625 45. Foreign entry from the perspective of an evolutionary theory of the firm.

Lieberman, M. B., and D. B. Montgomery. 1988. "First Mover Advantages." *Strategic Management Journal* 9 (Special Issue: Strategy Content Research): 41 58. A classic paper on entry timing.

Meyer, K. E., and Y.T.T. Tran. 2006. "Market Penetration and Acquisition Strategies for Emerging Economies." *Long Range Planning* 39 (2): 177 97. An applied paper illustrating practical aspects of developing foreign mar ket entries.

Welch, L. S., G.R.G. Benito, and B. Petersen. 2007. *Foreign Operation Methods: Analysis, Strategy, and Dynamics.* London: Edward Elgar. A new textbook.

Young, S., J. Hamill, C. Wheeler, and J. R. Davies. 1989. *International Market Entry and Development.* Hemel Hempstead, UK: Harvester Wheatsheaf. A classic, spe cialized textbook.

KLAUS MEYER

■ forward premium puzzle

Interest rates, the exchange rate, and the forward rate are all linked together by expectations and individual attempts to obtain the highest rates of return. The forward premium puzzle is closely related to the failure of uncovered interest parity (UIP) to hold, and the phenomenon of forward rate bias, which refers to the tendency for the forward exchange rate (the price of foreign currency agreed to today for a transaction some specific time in the future) to sys tematically mispredict the future spot exchange rate (the exchange rate in transactions in which one cur rency is directly exchanged for another). The puzzle is the finding that the forward premium the gap be tween the spot exchange rate and the forward rate usually points in the wrong direction for the subse quent actual movement in the spot exchange rate. Uncovered interest parity states that if the covered interest parity holds for a pair of currencies, then the forward discount, and hence the interest differential between the two countries, should be an unbiased predictor of the subsequent change in the spot rate, assuming that actors make guesses about the future

that are on average correct (what is sometimes termed "rational expectations"). The puzzle arises from the fact that, if market participants were risk neutral, the forward rate should equal the future expected spot exchange rate; yet when the forward rate suggests depreciation, typically the exchange rate appreciates, and vice versa.

The puzzle is of importance for what it suggests about the workings of international financial mar kets. To the extent that the puzzle reflects the failure of rational expectations to hold, then interest rate differentials will be poor guides to future exchange rate movements. In addition, the implied lack of market efficiency suggests a potential role for gov ernment intervention in foreign exchange markets. On the other hand, if the puzzle reflects the presence of a premium to compensate for the riskiness of specific currencies, then capital may not be so ready to migrate from one currency to another.

Explaining the Forward Premium Puzzle To fix concepts and terms, define the forward rate at time t for a trade to occur at time k as F_t^k and the spot rate at time t as S_t. Further, let the subjective expectation of the spot rate at time $t+k$, based on time t informa tion, be defined as $\varepsilon_t(S_{t+k})$. Assume for the moment rational expectations, namely, $E_t(S_{t+k})$. Then one should expect:

$$S_{t+k} = F_t^k + \tilde{u}_{t+k} \qquad (1)$$

where the error term is an expectational error.

In reality, regression estimates do not obtain a regression coefficient of unity, although the point estimate is often not statistically significantly far from the posited value.

The forward premium puzzle can be identified by assuming that the error term is log normally dis tributed, so that (1) can be rewritten as:

$$s_{t+k} = \beta_0 + \beta_1 f_t^k + \tilde{u}_{t+k} \qquad (2)$$

where under the null hypothesis, $\beta_1 = 1$, and β_0 is allowed to equal some constant impounding some Jensens Inequality terms.

Notice that one can subtract the current log spot rate s_t from both sides, since under the null $\beta_1 = 1$. This yields:

$$s_{t+k} - s_t = \beta_0 + \beta_1(f_t^k - s_t) + \tilde{u}_{t+k} \qquad (3)$$

The left hand side of equation (3) is ex post depreci-ation, while the term in the parentheses is the forward discount (or the inverse of the forward premium).

The puzzle is that estimates of β_1 are not only different from the value of unity, and statistically significantly, but also that the coefficient estimates are typically *negative*. This suggests that agents could make substantial profits by arbitraging. To be con-crete, individuals could borrow in the low interest currency and lend in the high interest currency, in a process termed "the carry trade." This well-known strategy can be highly profitable, although the profits are particularly vulnerable to sharp movements in the exchange rate.

This issue is linked up to uncovered interest parity in the following sense. If covered interest parity holds, then:

$$(f_t^k - s_t) = (i_t^k - i_t^{k*}) \qquad (4)$$

Substituting this no-arbitrage-profits condition into (3), one finds that (3) can be rewritten as:

$$s_{t+k} - s_t = \beta_0 + \beta_1(i_t^k - i_t^{k*}) + \tilde{u}_{t+k} \qquad (5)$$

The equation above is the regression used to test the joint null hypothesis of uncovered interest parity and rational expectations. The finding of a negative slope coefficient in equation (5) is equivalent to the finding of a negative slope coefficient in (3), for instances where covered interest parity holds.

Reasons behind the Puzzle A number of papers have investigated a wide variety of different econo-metric issues, but overall it appears that the negative slope coefficient cannot be entirely explained by the time series characteristics of the variables. Leaving aside these econometric issues, the forward premium puzzle might exist even when capital is perfectly mobile according to the covered interest parity cri-terion, either because of the invalidity of the rational expectations hypothesis or the existence of an ex-change risk premium.

Estimates of equation (5) using values for k that range up to one year typically reject the unbiasedness restriction on the slope parameter. For instance, the survey by Kenneth Froot and Robert Thaler (1990) finds an average estimate for β of 0.88. Menzie

Chinn and Guy Meredith (2004) document that this result holds for more recent periods extending up to 2000. They also show that the bias decreases at longer horizons.

It is important to recall that uncovered interest parity, properly defined as relating to expected de-preciation, is untestable. Estimation of the standard UIP regression equation relies on the rational ex-pectations methodology embodied in equation (1). Of course, reliance on the assumption of rational expectations is by no means uncontroversial. In a number of papers, Kenneth Froot and Jeffrey Frankel (1989) demonstrate that the standard tests for UIP yield radically different results when one uses survey-based measures of exchange rate depreciation. They find that most of the variation of the forward dis-count appears to be related to expected depreciation, rather than a time-varying risk premium, thereby lending credence to UIP.

Menzie Chinn and Jeffrey Frankel (1994) docu-ment the fact that it is difficult to reject UIP for a broader set of currencies, although there is some evidence of a risk premium at the 12-month horizon. The authors interpret the differing results as arising from a wider set of currencies they examine 17 currencies as opposed to the 5 or so examined by Jeffrey Frankel and Kenneth Froot (1987) where the assumption of perfect substitutability of debt instruments is less likely to hold. As these authors have emphasized, rejection of the rational expecta-tions hypothesis does not necessarily mean one ac-cepts the proposition that agents are irrational. It may be that agents are constantly learning about the economic environment such that their forecasts are biased for long stretches of time.

Perhaps the most natural explanation for why the forward premium predicts the wrong direction of exchange rate movements is that a risk premium drives a wedge between expected changes and actual changes. However, the modeling of the risk premium has proven quite challenging. A standard model motivates the risk premium as a function of the correlation between relative returns on assets denom-inated in the two currencies, and the ratio of mar-

ginal utilities of consumption in the two respective countries. While the theory is quite straightforward, an implausibly high degree of risk aversion is necessary to rationalize the observed volatility of the risk premium. Some recent work has had success relying on different utility functions. For instance, Moore and Roche (2002) assume that the amount of utility that consumers gain from consumption depends in part on the amount of consumption undertaken in the previous period. This approach rationalizes the observed volatility with the seemingly small changes in consumption.

See also capital mobility; carry trade; foreign exchange intervention; interest parity conditions; monetary policy rules; peso problem; sovereign risk; sterilization

FURTHER READING

Alvarez, F., A. Atkeson, and P. J. Kehoe. 2002. "Money, Interest Rates, and Exchange Rates in Endogenously Segmented Markets." *Journal of Political Economy* 110 (1): 73 112. Offers risk based explanations for the for ward premium puzzle.

Baillie, R. T., and T. Bollerslev. 2000. "The Forward Pre mium Anomaly Is Not as Bad as You Think." *Journal of International Money and Finance* 19 (4): 471 88. Argues that there is a nonlinearity in the relationship between the spot rate and the forward discount. When the for ward discount is large in absolute value, then the forward discount is likely to point in the right direction. When the forward discount is small, it is likely to point in the wrong direction, perhaps because transactions costs are large relative to potential gains.

Chinn, M. D., and J. A. Frankel. 1994. "Patterns in Ex change Rate Forecasts for 25 Currencies." *Journal of Money, Credit, and Banking* 26 (4) (November): 759 70. Assesses the characteristics of forecasts as measured by surveys of market participants.

Chinn, M. D., and G. Meredith. 2004. "Monetary Policy and Long Horizon Uncovered Interest Parity." *IMF Staff Papers 51* (3) (November): 409 30.

Engel, C. 1996. "The Forward Discount Anomaly and the Risk Premium: A Survey of Recent Evidence." *Journal of Empirical Finance* 3 (June): 123 92. Surveys the large literature concerned with modeling the risk premium.

. 1999. "On the Foreign Exchange Risk Premium in Sticky Price General Equilibrium Models." In *Inter national Finance and Financial Crises: Essays in Honor of Robert P. Flood*, edited by Peter Isard, Assaf Razin, and Andrew Rose. Norwell, MA: Kluwer and IMF, 71 85. Offers a risk based explanation for the forward puzzle.

Frankel, J. A., and C. M. Engel. 1984. "Do Asset Demands Optimize over the Mean and Variance of Returns? A Six Currency Test." *Journal of International Econo mics* 17 (3 4): 309 23. Examines the relationship between differential currency returns and the amounts of government bonds denominated in different currencies.

Frankel, J. A., and K. A. Froot. 1987. "Using Survey Data to Test Standard Propositions Regarding Exchange Rate Expectations." *American Economic Review* 77 (1): 133 53. Assesses the characteristics of forecasts as measured by surveys of market participants, and finds that these forecasts are typically biased.

Froot, K. A., and J. A. Frankel. 1989. "Forward Discount Bias: Is It an Exchange Risk Premium?" *Quarterly Journal of Economics* 104 (1): 139 61. Concludes that for the major currencies, the bias in forward rates is due to bias in exchange rate expectations.

Froot, K. A., and R. H. Thaler. 1990. "Foreign Exchange." *Journal of Economic Perspectives* 4 (3): 179 92. Surveys the literature on the forward discount.

Lewis, K. K. 1989. "Changing Beliefs and Systematic Ra tional Forecast Errors with Evidence from Foreign Ex change." *American Economic Review* 79 (4): 621 36. Incorporates Bayesian learning by agents about the economic environment such that their forecasts are bi ased for long stretches of time.

Lyons, Richard. 2001. *The Microstructure Approach to Ex change Rates*. Cambridge, MA: MIT Press. Appeals to institutional and microstructural factors to explain the presence of excess returns.

McCallum, B. T. 1994. "A Reconsideration of the Un covered Interest Parity Relationship." *Journal of Mone tary Economics* 33 (1): 105 132. Appeals to a mone tary reaction function that responds to exchange rate changes, thereby making interest rates endogenous in an economic sense.

Moore, M. J. 1994. "Testing for Unbiasedness in Forward Markets." *The Manchester School* 62 (supplement): 67 78. Provides a statistical interpretation for why the for ward rate is a biased predictor of the future spot exchange rate.

Moore, M. J. and M. J. Roche. 2002. "Volatile and Per sistent Real Exchange Rates without the Contrivance of Sticky Prices." IDEAS. Web site of the Department of Economics, University of Connecticut. Available at http://ideas.repec.org/p/may/mayecw/n1160402.html. Incorporates a utility function with "deep habits" to explain the gap between the forward and expected future spot exchange rates.

Villanueva, O. Miguel. 2005. "FX Dynamics, Limited Participation, and the Forward Bias Anomaly." *The Financial Review* 40: 67 93. Provides econometric re interpretations of McCallum (1994).

MENZIE D. CHINN

■ fragmentation

Fragmentation involves the decomposition of a production process into its constituent activities and the dispersion of those activities across national borders. It gives rise to cross-country production sharing and production networks. Under fragmentation, traded goods and services contain value added from more than one nation.

The ability to be "fragmented" into "production blocks," each capable of being carried out in a physically separate location, is in the first instance a matter of technology. Whether or not to take advantage of this facility, however, is an economic decision that depends on location-specific advantages, such as proximity to raw materials, factors of production, and markets. While the physical nature of production may remain largely unchanged, its organization undergoes significant transformation.

In addition to choosing locations for constituent activities, firms provide "service links" in order to facilitate communication and coordination and to allow for efficient and on-time transportation of parts and components within the network (Jones and Kierzkowski 1990). Hence, fragmentation will take place only where savings in production cost exceed the costs associated with service links.

When making their choices, firms must also determine whether to keep the dispersed activities under company management or to outsource them to independent contractors in so-called arm's-length relationships. The choice depends on economic as well as legal and institutional considerations; it involves trust, enforceability of contracts, and protection of intellectual property (Helpman 2006; Barba Navaretti and Venables 2004). A substantial proportion of cross-border production involves the foreign operations of the affiliates of multinational corporations, implying that a rising share of international trade is intrafirm trade.

Fragmentation, which has also been described as "breaking up the value-added chain," works particularly well between advanced and emerging economies and has grown rapidly in industries such as electronics, machinery, and textiles and apparel. Key factors facilitating this development have been reductions in trade restrictions and other barriers and cost-cutting innovations in transportation and communications technologies. The rapid spread of international production sharing is changing trade patterns, trade balance accounting, and the interaction between trade, the exchange rate, and other macroeconomic variables.

The spread of production sharing implies that traded products increasingly contain parts and components from more than one country. Hence domestic value added to a country's exports is often significantly less than the value of those exports. Chinese exports of electronic machinery, for example, incorporate imported components. Similarly, the imports of many countries contain components made in those countries. United States imports of automobiles from Mexico, for example, carry large amounts of U.S.-made components.

Production sharing gives rise to a new form of intraindustry trade, in which parts and components and finished products belonging to the same industry or sector pass back and forth among countries. This vertical intraindustry trade differs from the more

traditional horizontal intraindustry trade of different varieties of the same end product.

Analytical Issues The term *fragmentation* was first used in this context by Jones and Kierzkowski (1990). Alternative terminology includes intraproduct specialization and vertical specialization. The effects and implications of cross-border fragmentation may be examined from a variety of theoretical perspectives, including Ricardian, Heckscher-Ohlin (H-O), and imperfect competition (Arndt and Kierzkowski 2001). The first two are the workhorses of trade theory. In the Ricardian model, countries are differentiated by technology, so intraproduct specialization among countries participating in production sharing is determined by location-specific technological know-how. Countries with limited technological development will engage in low-tech production, while more advanced components and assembly will be produced in countries possessing the necessary technologies.

In the Heckscher-Ohlin framework, differences in factor endowments provide the basis for specialization, such that countries will specialize in activities that make intensive use of locally abundant factors of production. Labor-intensive component production and assembly will, therefore, be carried out in labor-abundant countries, other things being the same, while skill- and capital-intensive activities will be performed in more advanced countries. In that framework, fragmentation has effects on output, factor prices, trade, and welfare analogous to those associated with technological progress (Arndt 1997).

The factor-proportions view of vertical intra-industry specialization and trade is a key contribution of the H-O approach, which had been criticized in years past for its inability to explain horizontal specialization and trade. Indeed, it was this shortcoming that prompted the development of imperfect-competition models of international specialization and trade, in which consumers value variety and monopolistically competitive firms specialize in the production of variety. Intraindustry trade associated with production networks, on the other hand, consists of the movement across borders of parts, components, and end products belonging to the same industry.

Factor endowments, however, are not the only type of location advantage. Others include distance, border, country size, infrastructure, legal and financial systems, and regulatory and other policies. The effects of these variables on trade in general have been studied with the aid of gravity models, which posit that trade between two countries will fall with distance and the thickness of the border; that country size measured in various ways affects bilateral trade positively; and that strong and transparent legal and institutional structures, ample infrastructure, and protrade regulatory policies increase trade.

The gravity equation is now beginning to be employed in the context of production networks, for the purpose of which it must be amended in a variety of ways. There is, for example, more than one distance to be accounted for. In addition to the distance between the importer and exporter of the final product, there is distance between countries supplying parts and components and the country in which the final good is assembled. Regional clustering of supplier nations (as in Southeast Asia) is a distinct location advantage.

Foreign Direct Investment Production networks with developing economies typically require up-front capital formation, financed at least in part by inflows of foreign direct investment (FDI) and technology from advanced countries. Foreign investors erect production facilities and install infrastructure, bring in skilled workers, and transfer technology at levels of sophistication beyond those available in the country. While such activities violate several assumptions of the H-O model, that model is nevertheless useful in predicting the pattern of FDI flows.

The investment involved is vertical (VFDI), intended to support vertical or intraproduct specialization. In a world of cross-border production fragmentation, the H-O model predicts that VFDI will tend to flow from capital- and skill-abundant, labor-scarce countries to labor-abundant economies. The Ricardian model predicts that such flows will move from technologically advanced to technologically emerging economies.

As already noted, factor endowments are only one type of location advantage, and therefore cannot fully explain movements of FDI. Once again, the gravity model offers a way of accounting for additional location-specific considerations   including distance, country size, border effects, and various institutional and policy-related factors. The bulk of the existing empirical literature on FDI, however, covers periods dominated by horizontal foreign direct investment (HFDI), which prevails among advanced countries and which continues to be the dominant form of FDI.

According to this evidence, FDI responds positively to distance. The explanation lies in the fact that in horizontal specialization FDI is a substitute for exports. As distance and its costs increase, exports are replaced by on-site production for the local market and hence FDI flows rise. Vertical FDI, on the other hand, moves to support production sharing. Hence, when distance raises the costs associated with servicing a network, production sharing becomes unprofitable and FDI declines.

Multinational Companies While many multinationals go abroad to produce for the local market, an increasing number have become involved in production networks, a process in which affiliate operations play a key role. Often, the initial organizational structure follows a "hub-and-spoke" pattern, with affiliates in several countries trading directly with the parent company, but less with one another. As interactions become more sophisticated and infrastructure is developed, true networks evolve in which trade takes place among affiliates located in various countries, as well as between affiliates and local firms. This pattern is exemplified by Japanese multinationals operating in Southeast Asia (Kimura and Ando 2003).

Agglomeration of activities is important in this context, because it generates positive externalities. Clusters of multinationals operating in a given country or in several adjacent countries attract and train workers, improve access to finance and other services, and generate knowledge spillovers that raise productivity and reduce costs. The creation of such an environment benefits from supportive public policies and public investment in infrastructure. A classic example is Ireland and the electronics industry. Such an environment generates positive agglomeration effects, as workers with industry-specific skills migrate into the area and knowledge spillovers occur.

In this setting, local producers are encouraged to enter the industry as suppliers of goods and services to the multinationals. There is evidence that arm's-length trade is growing in Southeast Asia as foreign multinationals rely increasingly on local suppliers. For developing countries, this opens another path to industrialization.

Scale economies are important in production networks and are exploited by multinationals. Since neither the Ricardo nor Heckscher-Ohlin framework is capable of handling scale economies and externalities, the imperfect competition literature provides important insights. Scale economies come into play on both the production side and in the provision of service links among production blocks. If production of each part or component is concentrated in one location, for example, longer runs permit exploitation of internal scale economies at each location. Meanwhile, agglomeration and clustering of producers generate economies that are external to plant and firm.

In the provision of service links, the fixed costs associated with establishing a communications network can be spread over larger numbers of units and firms as the size of the network and the number of participating entities expand. Access to a service network encourages agglomeration of firms in a country or region. Where individual countries tend to be small, cooperative policies among countries are required to reap the benefits of agglomeration.

Productivity, Employment, and Wages The spread of fragmentation has sounded political alarm bells in advanced countries about jobs and wages not only in manufacturing, but in services industries, as "offshoring" of call centers, help desks, programming, and ticketing operations has proliferated. There is no doubt that jobs of workers whose functions are shifted to foreign locations are lost. But if cross-border sourcing of components raises competitiveness and lowers prices of final products or

services, thereby increasing sales and thus output, then jobs will be created elsewhere in the industry or more generally in the economy.

With respect to wages, there are two concerns. First, production sharing with low-wage countries may exert downward pressure not only on the wages of workers whose jobs are lost, but on wages generally in a "race to the bottom." Second, it may change the wage distribution against low-skilled workers and widen the wage gap between skilled and unskilled workers.

The available evidence is mixed, in part because it is not easy to separate the effects of fragmentation from those of technological change and of other factors affecting employment and wages (Feenstra 1998). The outcome further depends on the sector in which fragmentation occurs, on the skill ratios in the affected industries, and on the relationship of each skill category to capital. In a framework in which fragmentation occurs in many industries, moreover, there will be "onshoring" as well as "offshoring," so that jobs lost in one sector must be balanced against jobs gained in others. Under such circumstances, the problem is mainly one of matching unemployed workers with emerging jobs. This is the familiar trade-adjustment problem.

Regional Integration In recent years, regional production sharing and trade in parts and components has grown rapidly. This growth has been facilitated by preferential trade liberalization. While production sharing is generally welfare-improving under free trade and in preferential trade areas (PTAs), rules of origin that restrict the movement of components from outside the PTA can introduce significant elements of trade diversion.

Production sharing is changing the nature of regional integration, in that it requires deeper integration than the removal of tariffs and other restrictions on the flow of goods and services. It calls for liberalization of investment and for the cross-border movement of persons, as well as harmonization of technical standards, regulatory policies, and dispute settlement procedures. To the extent that cross-border integration of production at the level of industries reduces asymmetries and promotes conver-

gence of business cycles, it helps pave the way for greater regional monetary cooperation, including monetary union.

When industries become linked across borders, production sharing tends to reduce the sensitivity of trade flows to exchange rate changes. Consider a depreciation of the Mexican peso against the dollar, which raises the price of imports of U.S.-made components for incorporation into passenger vehicles. When these vehicles are exported to the United States, the dollar's appreciation has an offsetting effect. Hence trade flows associated with production sharing are less sensitive to exchange rate movements than other types of trade. This has implications for intraregional exchange-rate arrangements (Arndt and Huemer 2007).

Fragmentation and Interdependence Production sharing allows countries to reap the benefits of increased specialization. In the context of standard trade theory, intraproduct specialization pushes out the production possibility curve and thereby allows countries to reach higher consumption frontiers. In this sense, its effects are analogous to those of technological progress and factor accumulation.

But it also links countries more closely to their trading partners and thereby reduces policy autonomy and increases interdependence. It goes well beyond trade in promoting economic integration by linking industries at the level of production. Countries are more open and thus more exposed to external shocks. A group of countries operating in a production network will experience similar shocks. As more industries become involved in networks, cross-country asymmetries decline. Business cycles tend to converge. The pressures and incentives for policy cooperation increase.

Fragmentation is an important feature of globalization. Products become internationalized, national markets are more closely linked to those abroad, economywide behavior patterns converge, and the domain of overlapping policy interests expands.

See also agglomeration and foreign direct investment; economies of scale; gravity models; Heckscher-Ohlin

model; intrafirm trade; intraindustry trade; outsourcing/offshoring; Ricardian model; technology spillovers

FURTHER READING

Arndt, Sven W. 1997. "Globalization and the Open Economy." *North American Journal of Economics and Finance* 8 (1): 71 79. Derivation of key results of frag mentation in terms of the Heckscher Ohlin Model.

Arndt, Sven W., and Alexander Huemer. 2007. "Trade, Production Networks and the Exchange Rate." *Journal of Economic Asymmetries* 4 (1): 11 39. Examines the effect of global production networks on the sensitivity of trade flows to exchange rates.

Arndt, Sven W., and Henryk Kierzkowski, eds. 2001. *Fragmentation: New Production Patterns in the World Economy*. Oxford: Oxford University Press. A collection of papers dealing with various aspects of fragmentation.

Barba Navaretti, Giorgio, and Anthony J. Venables. 2004. *Multinational Firms in the World Economy*. Princeton, NJ: Princeton University Press. A comprehensive study of the FDI, multinationals, and trade.

Deardorff, Alan V. 2001. "Fragmentation in Simple Trade Models." *North American Journal of Economics and Fi nance* 12 (2): 121 37. Provides a more formal assess ment of the effects of fragmentation.

Feenstra, Robert C. 1998. "Integration of Trade and Dis integration of Production in the Global Economy." *Journal of Economic Perspectives* 12 (4): 31 50. Overview of theoretical and empirical issues related to fragmen tation.

Helpman, Elhanan. 2006. "Trade, FDI, and the Organi zation of Firms." NBER Working Paper No. 12091. Cambridge, MA: National Bureau of Economic Re search. A comprehensive overview of modeling multi nationals.

Jones, Ronald W., and Henryk Kierzkowski. 1990. "The Role of Services in Production and International Trade." In *The Political Economy of International Trade*, edited by R. W. Jones and Anne O. Krueger. Oxford: Basil Blackwell, 31 48. Discussion of the role of service links in production networks.

Kimura, Fukunari, and Mitsuyo Ando. 2003. "Fragmen tation and Agglomeration Matter: Japanese Multi nationals in Latin America and East Asia." *North American Journal of Economics and Finance* 14 (3): 287

317. An empirical study of Japanese MNCs operating in the two regions with a focus on behavioral differences.

SVEN W. ARNDT

■ free trade area

A free trade area (FTA) is created by an agreement among a group of countries to eliminate trade bar riers on most (if not all) goods between them. FTAs can be, and often are, regional and dictated by geo graphical considerations, such as the North Ameri can Free Trade Area, composed of the United States, Canada, and Mexico. They can also occur between countries that are far apart, as with the U.S.-Israel or U.S.-Singapore FTAs. Countries can belong to more than one FTA. The number of FTAs has increased rapidly since the early 1990s. For example, as of October 2007, the World Trade Organization (WTO) reported that approximately 110 notified (that is, officially existing) FTAs (excluding acces sions) were active at that time.

In contrast to a customs union (CU), which sets a common external tariff, FTAs do not necessarily equalize their members' tariffs on nonmembers. With differences in tariffs, rules of origin (ROOs) prevent trade in a product from going through the country with the lowest tariff on it and then being shipped within the FTA, since a good is eligible for zero tariffs in the FTA only if it originates there and ROOs specify the conditions required for origin to be granted.

There are large differences in the effects of an FTA with and without ROOs. In the absence of ROOs, an FTA results in large changes in trade flows as trade seeks the lowest tariff entry point into the FTA. Goods are then transshipped to their final destination in the FTA. Of course this results in large tariff rev enue transfer effects as this trade deflection transfers tariff revenue to the country with the lowest tariff entry point. As pointed out by Richardson (1995), this can result in a race to the bottom in setting tariffs. Moreover, in the presence of transshipment costs, such arbitrage may waste valuable resources: if such waste outweighs the positive effects of lower tariffs,

an FTA without ROOs may even reduce welfare overall.

In the presence of ROOs, however, simple transshipment is not possible. Nevertheless, some trade deflection may still be possible. By shipping domestic production to its FTA partners and meeting domestic demand via imports, the low-tariff country can still attract trade to its ports.

The Legal Basis of FTAs Signatories to the General Agreement on Trade and Tariffs (GATT) accord most-favored-nation (MFN) status to one another. This means that they cannot offer anyone else a more favorable (lower) tariff. How then can FTAs, which are by their very nature discriminatory, be legal? The answer lies in Article XXIV of GATT, which explicitly allows FTAs and CUs as exceptions to this rule as long as substantially all trade is free among members, and trade barriers with nonmembers is not higher on average after the FTA or CU.

The Welfare Effects of FTAs Although one thinks of FTAs as liberalizing of trade, and trade liberalization as welfare improving, neither may be true. If one pictures an FTA as only reducing the tariff on each good to the lowest tariff set by any of the FTA members, one might be inclined to think that such an FTA is welfare improving. Even this is not necessarily so, however. The theory of the second best says that welfare could fall with liberalization if not all distortions are removed, and consequently, even in such a stylized setting, an FTA may reduce welfare.

The usual argument showing the possibility of welfare loss is Viner's (1950). If an FTA is formed with the high-cost supplier, what is called trade diversion occurs. The country imports from its higher-cost FTA partner rather than the lower-cost country outside the FTA, so the gain in consumer surplus or welfare from lower tariffs and resulting lower consumer prices could easily fall short of the loss in tariff revenue. Trade-diverting customs unions could reduce welfare, whereas trade-creating customs unions, those formed with the lowest-cost supplier, would raise welfare.

FTAs need not be liberalizing, since if ROOs are hard enough to meet, they will raise costs by enough that no firm will find it worthwhile to meet them in order to obtain zero tariffs. In this case, FTAs do nothing. It is even worse if ROOs are strict and raise costs a good deal, but are still worth meeting. In this case, even if prices fall a little, the loss from the cost increase due to ROOs, which wastes resources, and the tariff revenue forgone, could easily reduce welfare relative to that prior to an FTA.

Moreover, recent work in this area has shown that the goals of improved market access and welfare may well be in conflict. Ju and Krishna (2000) show that if the excess demand for exported goods does not respond to changes in the prices of imported goods, then any policy that increases imports must also reduce welfare. Anderson and Neary (2007) interpret these results in terms of the generalized mean and variance of tariffs. They show that welfare is negatively related and import volume is positively related to the generalized variance and this causes a tension in the two objectives.

Stepping Stones, Stumbling Blocks, or Building Blocks? Of significant policy relevance is whether FTAs help or hurt the chances of further liberalization along multilateral lines. Bhagwati (1991) has been a vocal opponent of FTAs, arguing that they dampen the enthusiasm of a country for nondiscriminatory reform such as the multilateral trade negotiations held under the auspices of GATT/WTO.

The first question to ask is why multilateral free trade should be seen as more desirable than bilateral free trade. After all, if an FTA is a microcosm of the whole world, then there is not much to gain from global free trade. This argument neglects two factors, however. First, increasing the size of the world does raise welfare when goods are differentiated (as larger economies have greater variety and therefore higher real income and welfare) or there are increasing returns to scale (as larger economies can better exploit economies of scale and therefore have lower prices and higher welfare). Second, bilateral arrangements are not usually of this form but tend to occur between similar countries first, extending to others later. This makes sense, as such FTAs are likely to face less resistance as they have little effect on factor prices and

bring gains from reaping economies of scale and greater product variety.

Levy (1997) argues persuasively that, for political economy reasons, bilateral free-trade agreements can undermine political support for further multilateral trade liberalization. He uses a median voter model where agents have different endowments of capital, and where bilateral free trade could occur before multilateral free trade. With perfect foresight, no proposal that makes the median voter worse off than under multilateral free trade can occur. Thus if multilateral free trade is not possible directly, it cannot become so after a bilateral agreement. As a result, a bilateral agreement cannot be a stepping stone to multilateral free trade. It can, however, be a stumbling block. A free trade area has two effects in his model. It reduces the earnings of the relatively scarce factor in a country via standard Hecksher-Ohlin channels (since via trade, the relatively scarce factor in a country becomes in effect less scarce and so its price falls), and it raises the real income and, hence, welfare of all agents via increased variety. When the latter effect dominates, as it does if the FTA was among similar countries, a bilateral trade agreement makes most agents better off. But by doing so, it raises the reservation utility (the welfare under the status quo) for future multilateral agreements. Multilateral free trade would raise the gains to those with the most capital but reduce the gains to those with little capital, and as a result the median voter could be worse off with multilateral free trade. Thus, even if total welfare is highest under multilateral free trade, bilateral arrangements such as an FTA can prevent it from occurring.

A classic argument by Johnson (1967) is that multilateral reform would dilute existing preferences and therefore would provoke resistance by those adversely affected. For example, Mexican suppliers who are at an advantage under the North American Free Trade Agreement (NAFTA) would lobby against a reduction in the U.S. MFN tariff. Under the EBA (Everything but Arms) agreement of the European Union (EU), the poorest countries are allowed to export all goods other than arms to the EU market free of tariffs. If the EU reduced its MFN tariff, however, these preferences would be diluted and these poorest counties would lose their EU market to more competitive suppliers and would lobby hard against such liberalization on the part of the EU.

Limão (2006) has shown that there is evidence that such concerns are important. Using data on U.S. tariff reductions during the Uruguay Round of WTO negotiations, he shows that tariffs tended to be higher on goods that were traded in NAFTA. Moreover, as a result, the reciprocating tariff cuts tended also to be smaller.

Baldwin (2006) takes the opposing viewpoint. He argues that FTAs are building blocks for multilateral liberalization. He argues for a domino effect: once a group of countries form an FTA, then even if some countries stayed out of the FTA to begin with, once it is large enough, they would find themselves with so few trading partners that they would want to join! As a result, maybe there would be waves of integration: first a regional trading arrangement forms with a core group, then the next tier of countries joins, and so on. What happens more often, however, is that countries belong to many FTAs, and free trade is not a transitive relation. If A has an FTA with B and B has an FTA with C, this is not the same as A having an FTA with B and C. Goods from A will not have origin in B and therefore will not get zero tariffs when exported to C. Some exports from A via B to C could occur, however, with B's production being exported to C and B consuming A's products.

What, then, is the role of FTAs today and what can we expect in the future? Clearly, FTAs are here to stay. We cannot expect overlapping FTAs to do what multilateral free trade would, and there is empirical evidence that they act as stumbling blocks to global free trade. For these reasons, their proliferation is cause for concern. On the plus side, however, it is likely that their formation results in economic gains, both static and dynamic. Since FTAs may be possible (as sensitive sectors can be isolated via the use of restrictive ROOs) when multilateral free trade is not, it is as yet far from obvious what their net welfare effects are.

See also common market; customs unions; Free Trade Area of the Americas (FTAA); multilaterialism; North American Free Trade Agreement (NAFTA); regionalism; rules of origin

FURTHER READING

Anderson, James E., and Peter Neary. 2007. "Welfare versus Market Access: The Implications of Tariff Structure for Tariff Reform." *Journal of International Economics* 71 (1): 187 205. Explains why the mean and the variance of tariffs matter in different ways for welfare and market access.

Baldwin, Richard E. 2006. "Multilateralizing Regionalism: Spaghetti Bowls as Building Blocs on the Path to Global Free Trade." *World Economy* 29 (11): 1451 1518. Discusses the possible positive effects of FTAs.

Bhagwati, Jagdish. 1991. *The World Trading System at Risk.* Princeton, NJ: Princeton University Press. A nontechnical introduction to the role of the WTO in the global economy.

Johnson, Harry G. 1967. *Economic Policies towards Less Developed Countries.* Washington DC: Brookings Institution. Outdated, but a classic.

Ju, Jiandong, and Kala Krishna. 2000. "Welfare and Market Access Effects of Piecemeal Tariff Reform." *Journal of International Economics* 51 (2): 305 16. Explores the trade off between market access and welfare in the context of piecemeal policy reform.

Levy, Philip I. 1997. "A Political Economic Analysis of Free Trade Agreements." *American Economic Review* 87 (4): 506 19. An important early paper on the stepping stones versus stumbling blocks debate.

Limão, Nuno. 2006. "Preferential Trade Agreements as Stumbling Blocks for Multilateral Trade Liberalization: Evidence for the U.S." *American Economic Review* 96 (3): 896 914. Shows that the presence of an FTA seems to be correlated with reduced multilateral liberalization in goods traded in the FTA.

Richardson, Martin. 1995. "Tariff Revenue Competition in a Free Trade Area." *European Economic Review* 39: 1429 37. Looks at the incentives for tariff setting in an FTA.

Viner, Jacob. 1950. *The Customs Union Issue.* New York: Carnegie Endowment for International Peace. Old but still a classic.

KALA KRISHNA

■ Free Trade Area of the Americas (FTAA)

The United States and Latin America have long resisted what geography would seem to dictate: a special inter-American relationship built around open trade and investment flows. U.S. foreign policy has generally focused on other regions of the world notably Europe and Asia and eschewed regional favorites altogether in favor of a global reach. In Latin America, nationalists of the right and left have preferred to limit their dependence on U.S. power by diversifying their relations through stronger ties to Europe or other Latin American nations. Nevertheless, there have been periods in history when Latin America has reached out to the United States and the United States has responded affirmatively. The 1990s were one such period, and the centerpiece of such inter-American cooperation was the Free Trade Area of the Americas (FTAA).

The idea of a hemispheric free trade zone was not new when policymakers began to discuss it in earnest in the final decade of the 20th century. The goal can be traced back at least to Simón Bolívar (whose integrationist vision sometimes excluded but sometimes seemed to include North America), and it was discussed at the time of the founding of the Pan-American Union at the end of the 19th century. Presidents Ronald Reagan (1981 89) and George H. W. Bush (1989 93) made rhetorical references to the idea of a hemispheric free trade zone. However, this vague aspiration only became a concrete policy option when Latin American governments pressed a reluctant United States to take their free-trade agenda seriously. In 1990, it was Mexican president Carlos Salinas de Gortari who proposed a free trade accord to President Bush the seed that became the North American Free Trade Agreement (NAFTA). It was the Chileans who pressed three successive U.S. administrations for a free trade agreement. It was the Latin Americans who proposed to the Bill Clinton White House that the United States convene a post-NAFTA meeting of hemispheric leaders to spread the spirit of NAFTA southward, and who insisted that the centerpiece of the subsequent 1994 Miami Summit be the Free Trade Area of the Americas (FTAA).

The Latin Americans had good reasons to pursue hemispheric integration. An FTAA would permit them secure access to the world's largest and most dynamic market, and would signal to investors a more stable policy environment and a warmer and more predictable business climate. A regional free trade alliance would focus national policy debates on the next stage of reforms (following on the controversial but generally successful Washington consensus financial stabilization programs) required to promote international competitiveness, including further market liberalization, effective regulation and competition policies, and modern infrastructure in, for example, telecommunications, energy, and transportation. Politically, Latin America's democratic leaders were looking to Washington to bolster their position against the ever-present authoritarian tendencies in Latin American polities.

For the United States, an FTAA would open markets for trade and investment, giving U.S. firms preferential access against competitors from Europe and Asia. It could help Latin American countries to lock in macroeconomic reforms, accelerate steps toward market liberalization, and enhance the legitimacy of still-fragile democratic governments. Furthermore, on issues from counternarcotics to counterterrorism, environmental protection to energy cooperation, free trade was widely understood as generating positive externalities or spillover effects in diplomatic negotiations. Countries linked together in an FTAA could also serve as useful allies in global forums, including in World Trade Organization (WTO) trade negotiations. At the same time, fearing job loss especially in low-wage goods, environmental destruction, and a broad-based lowering of social standards, in a "race to the bottom," powerful, contrary voices in the United States spoke out vociferously against a regional trade accord and accelerating globalization more generally. Still other critics questioned the value of focusing trade negotiations on Latin America, as opposed to larger markets in Asia and Europe, and the WTO.

From 1994 to 2005, hemispheric trade negotiations proceeded on three tracks. As follow-on to the Miami Summit and the launch of the FTAA nego-

tiations, an elaborate web of negotiating committees and working groups evolved into a network of as many as 1,000 officials. But, with a deadline of 2005, negotiations proceeded at a leisurely pace and repeated draft texts left the more contentious issues in brackets. The United States pressed for liberalization in manufacturing and services, while Brazil and other agricultural exporters focused on U.S. protectionism and fiscal subsidies in products such as sugar, beef, cotton, peanuts, and orange juice that undercut the competitiveness of Latin American producers. In 2002, on the eve of off-year elections, the U.S. Congress passed a farm bill that sharply increased such agricultural subsidies, embarrassing U.S. trade officials who for years had been chastising the Europeans and Japanese for protecting their domestic farmers at the expense of global efficiency and developing-country producers.

The two leading countries in the FTAA negotiations, the United States and Brazil, turned to the Doha Development Round of negotiations within the framework of the WTO to help resolve some of the agricultural issues. However, the global WTO seemed no more capable of cutting the Gordian knot of agricultural subsidies than did the regional FTAA forum.

As the Doha Round prospects dimmed and the FTAA negotiations stumbled, many countries turned to less ambitious, more manageable bilateral and subregional trade arrangements. Following NAFTA, the United States ratified similar agreements with Chile and Central America (CAFTA), and concluded negotiations with Panama, Colombia, and Peru (approved by the Colombian and Peruvian legislatures but by the end of 2007 the U.S. Congress had approved only the Peruvian accord). Trade-oriented Mexico and Chile negotiated numerous bilateral accords with other Latin American nations as well as with the European Union and a growing number of Asian nations.

As the FTAA negotiations slowed and the 2005 deadline approached, neither the United States nor Brazil was willing to tackle domestic protectionist interests to close the deal. The Bush administration became embroiled in Iraq and its domestic leverage

declined, just as the Democrats in Congress became emboldened and, increasingly, fell under the influence of organized labor and other voices who demanded a new "fairer" approach to globalization. As the overall U.S. trade deficit ballooned, Democrats held proposed trade accords hostage, as leverage to overhaul other aspects of U.S. trade policy, and to better position the Democrats in upcoming elections. In Brazil, the 2002 election of Luiz Inácio Lula da Silva strengthened the nationalist forces in the powerful foreign ministry, Itamaraty, many of whose diplomats perceived the FTAA as strengthening the geopolitical position of the United States at the expense of Brazil's hegemonic aspirations in South America.

At the 2005 Summit of the Americas at Mar del Plata, Argentina, it became clear that political winds were shifting in South America. Argentine president and summit host Nestor Kirchner allied with Venezuelan president Hugo Chávez in adopting highly critical postures toward the FTAA. Emboldened by strong trade surpluses driven by the global commodity boom, the two outspoken Latin American leaders were critical of U.S. foreign policy and in particular President George W. Bush, attacking "neoliberalism" and the Washington consensus in favor of more state-directed, protectionist domestic economic policies. The 2001 2 crash of the Argentine financial system had embittered relations between Argentina and the United States and the International Monetary Fund (IMF), while the Kirchner administration focused more on domestic recovery than on international trade negotiations. In an overt challenge to the FTAA, Chávez launched his Bolivarian Alternative for the Americas (ALBA), and enlisted Cuba, Bolivia, and Nicaragua in what appeared to be mainly a series of foreign assistance and barter agreements, exchanging subsidized Venezuelan oil for trade and, in the case of Cuba, technical assistance. Kirchner helped Chávez pursue membership in the four-nation Mercosur in part to offset the dominance of Brazil within the South American integration scheme injecting that accord with a more ideological and anti-U.S. flavor.

Thus by 2007 the FTAA seemed to have reached a halfway point. A series of agreements links Canada, the United States, Mexico, the Central American nations, Chile and Peru, and possibly Colombia and Panama (pending U.S. congressional approval). NAFTA has facilitated dramatic increases in trade and investment flows and has generated a number of proposals to deepen the tripartite integration in areas such as infrastructure, energy, education, and immigration; politically, NAFTA weathered the challenge of the hotly contested 2006 presidential elections in Mexico. CAFTA has captured the imaginations of the Central American governments and private sectors; Sandinista leader Daniel Ortega, on regaining the Nicaraguan presidency in 2006, said his nation should take advantage of the market opportunities opened by CAFTA. The NAFTA and CAFTA nations faced the challenge of better integrating those accords (together with Chile and Peru, and possibly Colombia and Panama) to reduce inconsistencies and, possibly, create an umbrella governance mechanism. For its part, South America has become progressively fragmented by conflicting ideologies, economic policy preferences, and leadership styles, and none of the proposed integration schemes the FTAA, Mercosur, ALBA seems capable of bridging the divides. The promise, often voiced by Brazil, that Mercosur could serve as a stepping stone toward the larger FTAA, seems increasingly remote.

The future of the FTAA, therefore, remains uncertain. Has a historic opportunity been lost, or was the hemisphere simply experiencing a pause in the long march toward regional economic integration? One scenario, based on the coalescing of new domestic coalitions forming common cause against local vested interests, envisions a renewed push in the United States and Brazil to complete the deal. In such a scenario, other South American countries might want to avoid being left out and join the Brazilian-U.S. bandwagon, marginalizing Chávez and his ALBA. Alternatively, the hemisphere might become increasingly divided between a northern tier NAFTA plus CAFTA and a southern tier fractured by the simmering rivalries among Venezuela, Brazil,

and Argentina, and the separate path chosen by an increasingly open, globalized Chile.

See also Andean Community; Central American Common Market (CACM); Central American Dominican Republic Free Trade Area (CAFTA-DR); free trade area; International Monetary Fund (IMF); Mercosur; North American Free Trade Agreement (NAFTA); Washington consensus; World Trade Organization

FURTHER READING

Corrales, Javier, and Richard Feinberg. 1999. "Regimes of Cooperation in the Western Hemisphere: Power, In terests, and Intellectual Traditions." *International Stud ies Quarterly* 43 (March): 1 36. Traces the ups and downs of inter American cooperation in the last 200 years, finding that major leaps forward in inter Amer ican cooperation occur only when prointegration forces are strong in both the United States and leading Latin American countries. This oft cited article was published before the deterioration in inter American relations that transpired during the early 21st century.

Elliott, Kimberly Ann. 2004. *Labor Standards, Development, and CAFTA.* International Economics Policy Briefs 04 2. Washington, DC: Institute for International Eco nomics. A balanced review of the arguments for and against including labor standards in international trade agreements.

Feinberg, Richard. 2002. "Regionalism and Domestic Politics: U.S Latin American Trade Policy in the Bush Era." *Latin American Politics and Society* 44 (4): 127 51. Argues that Latin Americans have sought to impose their free trade policy agenda on a reluctant United States and assesses the trade policies of the Bush administration and the divisions within the U.S. political system regarding trade with developing countries.

Jatar, Ana Julia, and Sidney Weintraub, eds. 1997. *In tegrating the Hemisphere: Perspectives from Latin America and the Caribbean.* Washington, DC: The Inter Amer ican Dialogue. A valuable collection of essays by leading Latin American economists explaining the trade policies of key countries and subregions.

Salazar Xirinachs, Jose Manuel, and Maryse Roberts, eds. 2001. *Toward Free Trade in the Americas.* Washington, DC: Brookings Institution. Provides an overview of in tegration processes in the Americas at the turn of the millennium, identifies areas of convergence and diver gence among the various regional trade agreements, and explores in detail the complicated, unresolved issues that eventually stymied FTAA negotiators.

Schott, Jeffrey J. 2001. *Prospects for Free Trade in the Americas.* Washington, DC: Institute for International Economics. Explores the rationales for inter American economic integration and, while offering a generally optimistic assessment of prospects for the FTAA, also analyzes the serious obstacles to progress. Considers the debates and procedures in the U.S. Congress and the "readiness" of individual Latin American countries to participate in a hemispherewide free trade zone.

RICHARD FEINBERG

G

gains from trade

The theory of international trade can be viewed as an application of positive economics (economics focused on establishing facts, in contrast with normative economics, which focuses on questions of policy). This theory seeks to determine the prices at which commodities are exchanged on world markets (the so-called terms of trade) and traces the causes of trade to comparative advantage, the principle by which a nation that opens to international trade is led to specialize in and export certain commodities and to import others. From the theory's inception in the early 19th century, economists showed an even stronger interest in the gains from trade, that is, the welfare implications of trade for the countries that participate in it. As Richard E. Caves (1960, 6) points out, "Classical international trade theory was concerned first of all with the gains from trade, and only occasionally with the analysis of international trade by means of a formal, determinate model." Aside from formal modeling, it remains true that gains from trade and causes of trade were closely linked by economists of the classical school because of their strong interest in economic policy. Their main reason for studying the causes of trade is that it benefits countries that engage in it. The intimate connection between trade and its welfare implications is captured in the titles that Jacob Viner assigned to the last two chapters of his magisterial *Studies in the Theory of International Trade*: "Gains from Trade: The Doctrine of Comparative Costs" and "Gains from Trade: The Maximization of Real Income." As Viner states, "Recognition of its 'welfare analysis' orientation is essential to the understanding and appraisal of the classical doctrine" (1937, 437).

Does international trade always benefit the countries that participate in it? Some observers, both economists and statesmen, have expressed skepticism about the existence of "gains" on a national level and believe they can carry a negative sign, in which case they should be labeled "losses from trade." Advocates for or against greater freedom of trade have argued with each other for centuries, and there is no reason to expect such arguments ever to cease. Although it is important to recognize these controversies over the benefits of free trade to the world as a whole, to a subset of countries that contemplate the creation of a free trade area, or to a single country, they can be fully addressed only from a multidisciplinary perspective that includes economic, political, sociological, environmental, labor, and other effects.

The gains from trade are important for understanding the world economy since the perception of what such gains (or losses) consist of has led for centuries to the implementation of trade policies such as trade restrictions (tariffs or quotas), or conversely the liberalization of trade adopted in the forum of bilateral or multilateral negotiations. The term *gains from trade* can apply to various types of economic agents not limited to nation-states, such as regions of a country (economic or geographic) or even smaller units such as firms or individuals. Since nations are regarded in this entry as the main agents of interest, a question immediately arises: If some groups or individuals gain from trade (or from the partial removal of trade restrictions) while others

lose, does "the nation" gain or lose? This issue, which is considered in the next section, explains why trade and trade policies are such controversial issues at both the national and international level. Before examining how the gains from trade were described and quantified by pamphleteers, economists, and schools of economic thought since the early 18th century, it is helpful to set out how contemporary trade theorists define and measure them on the assumption that all individuals in a country are alike, or that any losers from trade are compensated.

Definitions of the Gains from Trade The gains from trade can be measured by either the *compensating variation* (CV) or the *equivalent variation* (EV). Their definitions given here can easily be generalized to economies with multiple commodities, but are best understood in the context of an economy with two commodities (good 1 and good 2). If the superscripts a and t refer to equilibrium values of variables in autarky (a term describing self-sufficiency or no trade) and free trade, the CV is the amount of income that would need to be withdrawn in an economy trading at the free trade prices (p_1^t, p_2^t) so that consumers can buy the same consumption bundle they chose under autarky, (c_1^a, c_2^a), rather than their preferred consumption bundle under free trade, (c_1^t, c_2^t). In algebraic terms,
$$CV = p_1^t c_1^t + p_2^t c_2^t - (p_1^t c_1^a + p_2^t c_2^a). \qquad (1)$$
The EV uses autarky prices rather than free trade prices as reference prices to measure the gain in welfare, so that
$$EV = p_1^a c_1^t + p_2^a c_2^t - (p_1^a c_1^a + p_2^a c_2^a). \qquad (2)$$
The EV can be interpreted as the amount of money that would have to be donated to consumers under autarky to allow them to purchase the same consumption bundle they choose under free trade. Since the consumption bundles (c_1^a, c_2^a) and (c_1^t, c_2^t) are valued at different relative prices in (1) and (2), the magnitudes of the CV and EV are usually different, and there is no reason to believe that either one is a superior measure of the gains from trade.

The CV and EV are illustrated in figure 1. If good 1 is plotted on the horizontal axis and good 2 on the vertical axis, DE is the concave production possibility frontier (PPF) that shows the output combinations

that a neoclassical economy can produce with its given technology and factor endowments. If A is the point of equilibrium in autarky, the autarky price ratio is given by the slope of p^a, a line tangent to DE at A. If trade becomes possible at terms of trade given by the slope of line p^t, the production point moves from A to P, and the consumption point from A to C. The economy's comparative advantage in good 1 leads it to specialize partially in that commodity, and to export BP of good 1 in exchange for CB of good 2. For this reason, CBP is known as the economy's "trade triangle." The exports of good 1, BP or JK, equal production OK minus consumption OJ, and the imports CB of good 2 equal consumption CJ minus production PK ($=$ BJ). If the exportable good 1 is the *numéraire*, or unit of account in which prices are measured, and AF is drawn parallel to p^t, the CV for the passage from autarky to free trade is FP. If CH is drawn parallel to p^a, the EV is GH, which usually differs in length from FP.

The CV and EV represent attempts to measure in terms of good 1 the "distance" between the consumption bundles C and A, or the enhanced welfare of consumers due to trade. A more intuitive way to recognize the existence of gains from trade is to note that the free trade consumption bundle C is located outside the PPF DE and is therefore unavailable to the economy in autarky (Corden 1984). Since C is chosen under free trade in preference to any point on
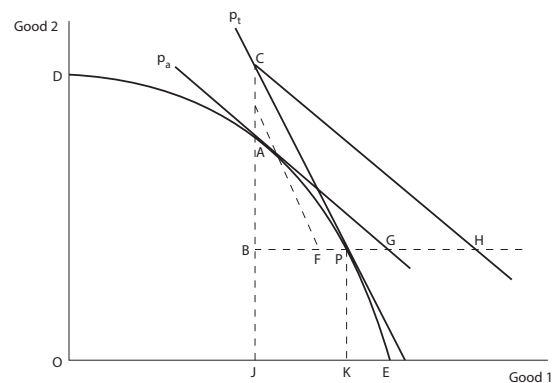
Figure 1
Gains from trade in a neoclassical economy

DE such as A, consumers are *potentially* better off under free trade since the gainers are always able to compensate the losers while remaining better off than in autarky (Samuelson 1939). In the absence of compensation, and in recognition of the fact that free trade usually produces both gainers and losers, a "social welfare function" is sometimes invoked that weights different persons' utilities according to the tastes of a dictator or decision-maker (benevolent or otherwise). The algebraic sum of these weighted utilities may be higher or lower than in autarky. Samuelson (1939) also made an observation, subsequently confirmed by other economists, that the gains from trade increase with the difference between the autarky and free trade price ratios. If these price ratios are identical, the economy remains in a state of autarky and there are no gains from trade.

If an economy trades according to its comparative advantage, does it always gain from trade? Leaving aside distributional issues, cases of foreign or colonial exploitation, "unequal exchange" and various distortions to which an economy may be subject (which are considered later), the answer is yes in the vast majority of cases. But there are exceptions where welfare under free trade is the same as in autarky, such as the case of an economy whose comparative advantage leads it to export a commodity it was not consuming before trade (Maneschi 1998a). Moreover when a large economy with a linear PPF attempts to trade with a much smaller one, it cannot fully specialize in its export commodity and faces terms of trade identical to its autarky price ratio, with zero gains from trade.

18th-Century Perspectives on the Gains from Trade In the 18th century, economists measured the gains from trade in accordance with what Viner calls the "18th-century rule," which postulates that "it pays to import commodities from abroad whenever they can be obtained in exchange for exports at a smaller real cost than their production at home would entail" (Viner 1937, 440), thus reflecting the benefits of trade as an indirect method of production. The earliest numerical example of the application of this "rule" was given by Henry Martyn (1701, 583): "If nine cannot produce above three Bushels of Wheat in *England*, if by equal Labour they might procure nine Bushels from another Country, to imploy these in agriculture at home, is to imploy nine to do no more work than might be done as well by three; . . . is the loss of six Bushels of Wheat; is therefore the loss of so much value." His numbers can be interpreted in terms of the two-commodity example of the previous section. In response to a cheaper overseas price of wheat (good 2), or terms of trade $(p_1/p_2)^t$ that exceed the autarky price ratio $(p_1/p_2)^a$, the economy tends to specialize in good 1. If $m_2 (= 9)$ bushels of wheat (good 2) are imported in exchange for x_1 units of good 1, balance of trade equilibrium requires that $m_2 = (p_1/p_2)^t x_1$. In autarky, x_1 units of good 1 are valued at $(p_1/p_2)^a x_1$, or 3 bushels of wheat. For every nine workers employed, the gains from trade (or *GT*) in terms of wheat are therefore

$$GT = m_2 - (p_1/p_2)^a x_1$$
$$= [(p_1/p_2)^t - (p_1/p_2)^a]x_1 \qquad (3)$$
$$= 9 - 3 = 6 \text{ bushels.}$$

Martyn's numerical example is a rudimentary attempt to quantify the equivalent variation EV defined earlier, if it is expressed in terms of the importable good 2. If the economy is in a state of autarky, it would have to receive foreign aid amounting to 6 bushels of wheat for every nine of its workers in order for them to reach the same welfare level they would enjoy under free trade.

The 18th-century rule was used to characterize the gains from trade by writers such as the French economist and statesman A.R.J. Turgot and the author of *The Wealth of Nations*, Adam Smith. Smith applied it to trade between tribesmen, between town and countryside, and between two nations. For example, laying great emphasis on the importance of trade between town and countryside, Smith noted that "the inhabitants of the country purchase of the town a greater quantity of manufactured goods, with the produce of a much smaller quantity of their own labour, than they must have employed had they attempted to prepare them themselves" (1776, 376). Since Smith never attempted to illustrate his examples with numbers, Martyn's 1701 numerical

illustration of the gains from trade stands out as a major analytical advance that was not equaled for more than a century, not even by someone as renowned as Smith (Maneschi 2002a).

The gains from trade were viewed in broader terms by the main French and British writers of the Enlightenment. Montesquieu and Voltaire praised commercial activity and luxury goods for their civilizing effects (*doux commerce*) and for promoting the power and military preparedness of nation-states. Smith's mentor and friend David Hume adopted similar views, observing that "commerce with strangers . . . rouses men from their indolence. . . . Imitation soon diffuses all those arts; while domestic manufactures emulate the foreign in their improvements, and work up every home commodity to the utmost perfection of which it is susceptible" (1752, 14). He later argued that "the encrease of riches and commerce in any one nation, instead of hurting, commonly promotes the riches and commerce of all its neighbours. Every improvement, which we have since [two centuries ago] made, has arisen from our imitation of foreigners. . . . Notwithstanding the advanced state of our manufactures, we daily adopt, in every art, the inventions and improvements of our neighbours" (1758, 78 79). Economists continue to maintain that trade encourages the international diffusion of technology, which may be one of its chief benefits. An international demonstration effect stimulates consumer demand for new commodities developed overseas and induces domestic firms to produce them. Hume also argued that the prosperity of neighboring nations promotes domestic prosperity instead of endangering it, as mercantilist writers had maintained.

In addition to using the 18th-century rule in various contexts, Adam Smith also took a broad perspective on the benefits of trade for producers as well as consumers (Maneschi 1998b). Consumers benefited not only from cheaper imports of staples but from the introduction of hitherto unknown commodities from Asia and the Americas. Europe's burgeoning market for manufactured exports allowed it greater scope for the division of labor and hence rising labor productivity. Another benefit that Smith emphasized was subsequently criticized by most classical economists: that trade offered a vent for surplus commodities produced over and above domestic needs. John Stuart Mill (1848, 579) called it "a surviving relic of the Mercantile Theory," since it implies that resources cannot be reallocated so as to avoid the "surplus" commodities whose existence Smith postulates. Smith also echoed Hume and French Enlightenment thinkers in praising trade for promoting economic interdependence. After arguing that "the discovery of America, and that of a passage to the East Indies by the Cape of Good Hope, are the two greatest and most important events recorded in the history of mankind," Smith observed that "by uniting . . . the most distant parts of the world, by enabling them to relieve one another's wants, to increase one another's enjoyments, and to encourage one another's industry, their general tendency would seem to be beneficial" (Smith 1776, 626).

David Ricardo's Twofold Gains from Trade In chapter 7 of his *Principles of Political Economy and Taxation*, Ricardo (1817, 132) claimed that international trade yields two distinct and coequal types of gain: "It is quite as important to the happiness of mankind, that our enjoyments should be increased by the better distribution of labour, by each country producing those commodities for which by its situation, its climate, and its other natural or artificial advantages, it is adapted, and by their exchanging them for the commodities of other countries, as that they should be augmented by a rise in the rate of profits." The first gain, the efficiency advantage arising from the reallocation of resources in accordance with comparative advantage, is unrelated to the second, a rise in the profit rate.

Ricardo quantified the first type of gain for two trading countries, England and Portugal, when he specified the amounts of labor embodied in each country's exports and the labor it would require to produce its imports:

England may be so circumstanced, that to produce the cloth may require the labour of 100 men for one year; and if she attempted to make the wine, it might require the labour of

120 men for the same time. England would therefore find it her interest to import wine, and to purchase it by the exportation of cloth. To produce the wine in Portugal, might require only the labour of 80 men for one year, and to produce the cloth in the same country, might require the labour of 90 men for the same time. It would therefore be advantageous for her to export wine in exchange for cloth. (Ricardo 1817, 135)

England thus saves 120 − 100 = 20 men by importing wine instead of producing it, and Portugal saves 90 − 80 = 10 men by importing cloth instead of producing it (Maneschi 2004). These gains from trade in terms of the labor each country economizes are another example of the "18th-century rule" cited earlier.

The second type of trade gain, incommensurable with the first, represents the increase in the rate of profit that trade may bring about: "It has been my endeavour to shew throughout this work, that the rate of profits can never be increased but by a fall in wages, and that there can be no permanent fall of wages but in consequence of a fall of the necessaries on which wages are expended. If, therefore, by the extension of foreign trade, or by improvements in machinery, the food and necessaries of the labourer can be brought to market at a reduced price, profits will rise" (Ricardo 1817, 132). This second type of gain depends exclusively on the import of commodities consumed by workers and is unaffected by that of luxury goods that do not enter the workers' consumer basket.

The increase in the profit rate achieved by trade was important for Ricardo since he had argued in previous chapters of the *Principles* that the pressure of population on land leads to diminishing returns at both the extensive and intensive margins of cultivation, to an increase in the price of "corn" (grains), to a rise in the money wage, and hence to a fall in the rates of profit and of capital accumulation. In the absence of trade or of technical change, the economy heads toward the stationary state. A trade-induced rise in the profit rate, resulting from the fall in the money wage when cheaper wheat is imported, reverses this trend by boosting the rate of capital accumulation. This "dynamic" gain from trade is additional to the "static" gain resulting from an improved resource allocation. Ricardo and the classical school thus closely associated trade with economic growth, and the gains from trade include those that affect an economy's trajectory over time. To judge from the frequency with which Ricardo referred throughout the *Principles* to the need for England to repeal the Corn Laws that held the price of grain artificially high, this second gain from trade was for him just as important as the saving of labor associated with the principle of comparative advantage (Maneschi 1992).

John Stuart Mill: Splitting the Gains from Trade When Ricardo postulated a trade equilibrium between England and Portugal, he stated each country's gains from trade but never explained how the terms of trade are determined. This task fell on the capable shoulders of John Stuart Mill, the son of James Mill, who had been one of Ricardo's closest supporters and disciples. John Stuart Mill pointed to reciprocal demand, the demand of each country for its trading partner's exports, as the force that determines the terms at which they trade (which he called the "terms of interchange"). His theory was first published in "Of the Laws of Interchange between Nations; and the Distribution of the Gains of Commerce among the Countries of the Commercial World," the first of his *Essays on Some Unsettled Questions of Political Economy* of 1844, and later in his *Principles of Political Economy* first published in 1848, which became the premier textbook of political economy in Great Britain. Mill argued that the gap between two countries' autarky price ratios, which reflect their respective comparative advantages, yields the range within which the terms of trade can settle, and that reciprocal demand determines where they settle within that range and hence how the gains from trade are split between them. According to Mill, "The produce of a country exchanges for the produce of other countries, at such values as are required in order that the whole of her exports may exactly pay for the whole of her imports. . . . So that supply and demand are but another expression for reciprocal

demand: and to say that value will adjust itself so as to equalize demand with supply, is in fact to say that it will adjust itself so as to equalize the demand on one side with the demand on the other" (Mill 1871: 592 93).

The terms of trade favor a country more the closer they lie to the *other* country's autarky price ratio. If the terms of trade coincide with the country's own autarky price ratio, its gains from trade are zero and all gains go to its trading partner. As a corollary of this, Mill observed that large countries gain less from trade than smaller ones "since, having a greater demand for commodities generally, they are likely to have a greater demand for foreign commodities, and thus modify the terms of interchange to their own disadvantage" (1871, 604). A later neoclassical economist, Francis Edgeworth, referred to this finding as "Mill's paradox."

In addition to Mill, other classical economists such as Robert Torrens emphasized the role of reciprocal demand in yielding the terms of trade. Torrens went even further than Mill. Since a foreign country's tariff can turn the terms of trade and the gains from trade in its favor, Torrens demanded reciprocity in trade policy and advocated a defensive tariff. Other economists criticized him for abandoning his previous policy in favor of unilateral free trade (O'Brien 2004). Torrens's important insight gave rise to Francis Edgeworth's argument, expressed more precisely later by Charles Bickerdike, that a "large" economy (one that can affect its terms of trade) maximizes its gains from trade by levying an optimal tariff.

In another chapter of his *Principles*, Mill explored further the nature of the gains from trade by pointing to its "indirect effects, which must be counted as benefits of a high order," including "the tendency of every extension of the market to improve the processes of production." Foreign trade creates a demonstration effect akin to an "industrial revolution" by inducing people to work harder so they can import commodities they were previously not familiar with: "The opening of a foreign trade, by making them acquainted with new objects, or tempting them by the easier acquisition of things which they had not

previously thought attainable, sometimes works a sort of industrial revolution in a country whose resources were previously undeveloped for want of energy and ambition in the people." Moreover, "the economical advantages of commerce are surpassed in importance by those of its effects which are intellectual and moral. It is hardly possible to overrate the value . . . of placing human beings in contact with persons dissimilar to themselves, and with modes of thought and action unlike those with which they are familiar" (1871, 581). For Mill, as for other classical economists such as Smith and Ricardo, the dynamic benefits of trade ranked at least as high as the static benefits emphasized by the neoclassical school that followed them (Maneschi 1998c).

Mill is also remembered as the first orthodox classical economist who, breaking rank with his confreres, advocated the infant-industry argument for protection in carefully restricted circumstances. A "young and rising nation" (he may have had the United States or Germany in mind) should be willing to forgo some of its current gains from trade by protecting a new industry in which other countries have a "present superiority of acquired skill and experience. A country which has this skill and experience yet to acquire, may in other respects be better adapted to the production than those which were earlier in the field" (Mill 1871, 922). In other words, a country may have a *potential* comparative advantage in an industry, but this potential needs to be fostered by temporary protection for it to be realized (other advocates of the "creation" of comparative advantage starting with Alexander Hamilton are discussed in Maneschi 1998b, chapter 5). This policy does not contradict free trade as a policy goal once an industry has matured and can compete in world markets, since protection then can and should be removed. Whether countries have actually followed this exemplary behavior is of course another question.

Neoclassical Depiction of the Gains from Trade
The classical theory of comparative advantage was attacked as based on an incorrect theory of value by economists of the neoclassical school that emerged in the last quarter of the 19th century. The Austrian-

born Gottfried Haberler (1936) argued that the prices of commodities are based not on the classical labor theory of value, but on their opportunity costs as measured by the least-cost combination of the factors that produce them. Diagrammatically he represented them by the slope of a concave production possibility frontier (PPF) such as that shown in figure 1. On this basis, he generalized the Ricardian theory of comparative advantage while retaining some its main policy implications.

Haberler was unaware that, in a long footnote inconspicuously embedded in a textbook of economic principles, the Italian economist Enrico Barone (1908, 88 90) had used a nonlinear PPF to depict in general equilibrium terms the gains accruing to an economy as it passes from autarky to free trade. He thus anticipated similar diagrams presented some 25 years later by Viner, Abba Lerner, and Wassily Leontief (Maneschi and Thweatt 1987). Barone's diagram is shown in figure 2, where AB is the PPF showing all possible combinations of commodities A (on the vertical axis) and B (on the horizontal) that the economy can produce. The dashed curves, now known as community indifference curves (CICs), show the combinations of the two commodities that yield the same societal welfare level. Autarky equilibrium is at M, where the CIC curve 3 is tangent to the PPF. The autarky price ratio is given by the common tangent at M (not shown) to both the PPF and the CIC. If the economy is given the opportunity to trade at terms shown by the slope of the line PC, with a higher relative price of B, it tends to specialize in B. The production point moves from M to P, the consumption point from M to C. The trade triangle CQP shows that exports QP of B exchange for imports CQ of A. The CIC labeled "8," which the economy can attain in free trade and is tangent to CP at C, clearly yields a higher welfare than under autarky, and illustrates the gains from trade.

The Heckscher-Ohlin-Samuelson Theory and Beyond In the 20th century the theory of international trade was revolutionized by Swedish economic historian Eli Heckscher (1949) and his student Bertil Ohlin (1933), who traced the sources of comparative
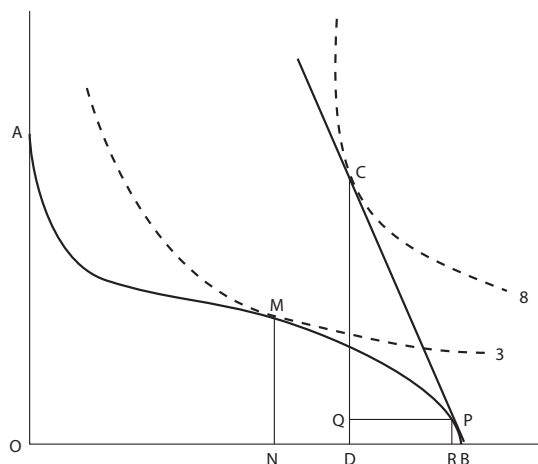
Figure 2
Barone's 1908 representation of the gains from trade

advantage to differences in relative factor endowments across countries and analyzed the effects of trade on income distribution. While Heckscher's intentions were reformist in nature, in the sense of modifying but not rejecting inherited trade theory, Ohlin aimed to overthrow the dated Ricardian theory and replace it with an explicitly neoclassical theory based on factor endowments. Both Heckscher and Ohlin noted that trade encourages the convergence of factor prices in the trading countries or regions. Ohlin, however, doubted that they could actually be equated and denied the possibility (as Haberler did before him) that a factor such as labor could be hurt by trade even if it is intensively employed in the import-competing industries whose prices are reduced by trade.

In a pathbreaking article, Wolfgang Stolper and Paul Samuelson (1941) showed that both Haberler and Ohlin were mistaken, since trade in fact can reduce the real wage of a factor such as labor if it is intensively used in the import-competing sector of the economy. Stolper and Samuelson proved this by stripping the Heckscher-Ohlin model down to two countries, two commodities, and two factors of production, thus obtaining results impossible to discern in the multicommodity, multifactor context of Heckscher and Ohlin. Samuelson (1948)

complemented this paper by setting out the stripped-down Heckscher-Ohlin model in diagrammatic form and showing how factor endowments and factor intensities affect comparative advantage and the pattern of trade. He proved that factor prices are fully equalized by trade in specific circumstances. The Stolper-Samuelson and factor price equalization theorems are components of what became the Heckscher-Ohlin-Samuelson theory of trade, which revealed the ambiguity in the expression *gains from trade* when applied to a country: if some factors gain while others lose from trade, as the Stolper-Samuelson theorem predicts must happen, how can the "nation" be said to gain? Samuelson showed that it can do so since the gainers can compensate the losers and still be better off than in autarky. While this is an important finding, it is cold comfort to factors that lose from free trade if in reality they are not compensated.

All the models of trade discussed so far postulate perfect competition in product and factor markets. Their conclusions and associated gains from trade are modified if this assumption does not hold. After adopting the Heckscher-Ohlin-Samuelson model as the mainstream theory of trade after World War II, economists analyzed how various types of distortions in factor or commodity markets affect the direction of trade and the welfare conclusions drawn earlier, and how tax or trade policies can be designed to alleviate these distortions. Policies can sometimes be ranked in terms of first-best, second-best, and so on (Bhagwati 1971; Corden 1997). In the absence of offsetting policies, it can be shown that autarky may be preferable to free trade. Economic growth can be immiserizing (Bhagwati 1958) by causing the terms of trade to deteriorate to such an extent that consumers are worse off than before growth. An optimal tariff, however, can avert this outcome. Trade policy is thus crucial in determining the existence and size of the gains from trade.

Starting in the 1970s, economists developed a New Trade Theory based on imperfect or monopolistic competition or oligopoly (see, for example, Krugman 1990). In the models they developed, comparative advantage is dispensed with as the source of trade, its place taken by economies of scale. Even countries identical in all respects can be shown to gain from trade if they specialize in different manufactures and reap the associated economies of scale. The nature of the gains from trade was substantially enriched by the New Trade Theory, since consumers are shown to benefit not only from cheaper commodities but from a greater variety of them than is available in autarky. Moreover trade can provide a dynamic impetus by stimulating innovation, technical change, and learning by doing. Dynamic benefits from trade that include the availability of new commodities and new methods of production are reminiscent of those identified more than a century earlier by classical economists such as Hume, Smith, and Mill (Maneschi 2002b). Even Ohlin (1933) devoted chapter 3 of his book to economies of scale as a secondary cause of trade in addition to factor endowments. The "New" Trade Theory can thus boast of numerous distinguished precursors.

The gains from trade have figured prominently in discussions of the rationale for international trade since mercantilist times and have typically led economists to argue for trade liberalization except in cases where they wished to promote infant industries or exploit monopoly power by turning the terms of trade in a country's favor. Attempts to quantify the gains from trade began in 1701 with Henry Martyn and continued in 1817 with David Ricardo and in the 20th century with the calculation of the compensating and equivalent variations. Classical economists such as David Hume and John Stuart Mill also argued for a broader vision of the gains from trade that includes the dynamic benefits that can be realized from the imitation of foreign ideas and technical improvements. Beginning in the 1970s, the exponents of the New Trade Theory added the availability to consumers of a greater variety of commodities as another important gain from trade.

See also absolute advantage; comparative advantage; economies of scale; Heckscher-Ohlin model; infant industry argument; intraindustry trade; monopolistic competition; New Trade Theory; Ricardian model; terms of trade

FURTHER READING

Barone, Enrico. 1908. *Principi di economia politica*. Rome: G. Bertero. Contains the first diagram illustrating the gains from trade in a neoclassical economy.

Bhagwati, Jagdish N. 1958. "Immiserizing Growth: A Geometrical Note." *Review of Economic Studies* 25 (3): 201 5. An often cited article proving that economic growth in a large country practicing free trade can make it worse off.

——. 1971. "The Generalized Theory of Distortions and Welfare." In *Trade, Balance of Payments, and Growth: Papers in International Economics in Honor of Charles P. Kindleberger*, edited by Jagdish N. Bhagwati et al. Amsterdam: North Holland, 69 90. An important analysis of the impact on economic welfare of various economic distortions in open economies and a ranking of the policies that can offset them.

Caves, Richard E. 1960. *Trade and Economic Structure: Models and Methods*. Cambridge, MA: Harvard University Press. A comprehensive exposition of models of international trade and their welfare implications.

Corden, W. Max. 1984. "The Normative Theory of International Trade." In *Handbook of International Economics*, vol. I, edited by Ronald W. Jones and Peter B. Kenen. Amsterdam: North Holland, 63 130. A noted trade economist surveys the welfare implications of international trade and of commercial policies.

——. 1997. *Trade Policy and Economic Welfare*. 2d ed. Oxford: Oxford University Press. An updated masterly survey of the welfare implications of trade and protection.

Haberler, Gottfried. 1936. *The Theory of International Trade*. Edinburgh: William Hodge. A comprehensive outline of neoclassical international trade theory and policy by one of its chief contributors.

Heckscher, Eli F. 1949. "The Effect of Foreign Trade on the Distribution of Income." In *Readings in the Theory of International Trade*, edited by H. S. Ellis and L. A. Metzler. Homewood, IL: Irwin, 272 300. First published in Swedish in 1919, the original and remarkably sophisticated presentation of the Heckscher Ohlin theory of trade.

Hume, David. 1752. "Of Commerce." In *David Hume: Writings on Economics*, edited by E. Rotwein. Madison: University of Wisconsin Press, 1955, 3 18. An essay that stresses the importance of foreign trade for the power of the state and the happiness of its citizens.

——. 1758. "Of the Jealousy of Trade." In *David Hume: Writings on Economics*, edited by E. Rotwein. Madison: University of Wisconsin Press, 1955, 78 82. In the last of his economic essays, Hume emphasizes that free trade combined with openness to foreign ideas and technical improvements enhances economic prosperity.

Krugman, Paul R. 1990. *Rethinking International Trade*. Cambridge, MA: MIT Press. An anthology of articles based on the new trade theory by one of its most eminent exponents.

Maneschi, Andrea. 1992. "Ricardo's International Trade Theory: Beyond the Comparative Cost Example." *Cambridge Journal of Economics* 16 (4): 421 37. Although Ricardo was a pioneer of the theory of comparative advantage, he arguably placed greater emphasis on the dynamic benefits of foreign trade in raising the rates of profit and capital accumulation.

——. 1998a. "Comparative Advantage with and without Gains from Trade." *Review of International Economics* 6 (1): 120 28. Illustrates the fact that gains from trade are sufficient but not necessary for the existence of comparative advantage.

——. 1998b. *Comparative Advantage in International Trade: A Historical Perspective*. Cheltenham: Edward Elgar. An analysis of the genesis, evolution, and current significance of the concept of comparative advantage in the theory of international trade.

——. 1998c. "The Dynamic Nature of Comparative Advantage and of the Gains from Trade in Classical Economics." *Journal of the History of Economic Thought* 20 (2): 133 44. Argues that classical economists regarded comparative advantage as subject to change, and the gains from trade as dynamic in nature rather than static.

——. 2002a. "The Tercentenary of Henry Martyn's *Considerations upon the East India Trade*." *Journal of the History of Economic Thought* 24 (2): 233 49. A celebration of the tercentenary of Martyn's pamphlet of 1701 that highlights his quantitative measure of the gains from trade.

——. 2002b. "How New Is the 'New Trade Theory' of the Past Two Decades?" In *Is There Progress in Economics? Knowledge, Truth, and the History of Economic Thought*,

edited by S. Boehm et al. Cheltenham: Edward Elgar, 240 55. The new trade theory that emerged in the 1970s had remarkable anticipations in the classical school and in Ohlin's (1933) book.

———. 2004. "The True Meaning of David Ricardo's Four Magic Numbers." *Journal of International Economics* 62 (2): 433 43. A challenge to the conventional interpretation of the numerical example that Ricardo used to illustrate comparative advantage.

Maneschi, Andrea, and William O. Thweatt. 1987. "Barone's 1908 Representation of an Economy's Trade Equilibrium and the Gains from Trade." *Journal of International Economics* 22 (3/4): 375 82. Shows that the first diagram illustrating the gains from trade in a neoclassical economy was published by Enrico Barone in 1908.

Martyn, Henry. 1701. *Considerations on the East India Trade*. In *Early English Tracts on Commerce*, edited by John R. McCulloch. Cambridge: Cambridge University Press, 1954, 541 629. The "reform mercantilist" Martyn's pamphlet supporting free trade and highlighting its gains.

Mill, John Stuart. 1871. *Principles of Political Economy*. Reprint of 7th ed. London: Longman, Green, 1920. First published in 1848, this became the most popular 19th century textbook of British classical political economy.

O'Brien, Denis P. 2004. *The Classical Economists Revisited*. Princeton: Princeton University Press. A thorough analytical survey of the contributions of the classical school of economics.

Ohlin, Bertil. 1933. *Interregional and International Trade*. Rev. ed. Cambridge, MA: Harvard University Press, 1967. A comprehensive and literary elaboration of the Heckscher Ohlin theory of trade.

Ricardo, David. 1817. *On the Principles of Political Economy and Taxation*. In *The Works and Correspondence of David Ricardo*, vol. I, edited by Piero Sraffa. Cambridge: Cambridge University Press, 1951. A fundamental contribution to classical economic theory containing a numerical illustration of comparative advantage and the gains from trade.

Samuelson, Paul A. 1939. "The Gains from International Trade." In *Readings in the Theory of International Trade*, edited by H. S. Ellis and L. A. Metzler. Homewood, IL: Irwin, 1949, 239 52. The first theoretical analysis of the gains from trade that can be realized by a country facing given world commodity prices.

———. 1948. "International Trade and the Equalization of Factor Prices." *Economic Journal* 58 (230): 163 84. Expounds and illustrates the Heckscher Ohlin Samuelson theory of trade.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Oxford: Clarendon Press, 1976. The most famous book of the founder of classical political economy.

Stolper, Wolfgang F., and Paul A. Samuelson. 1941. "Protection and Real Wages." In *Readings in the Theory of International Trade*, edited by H. S. Ellis and L. A. Metzler. Homewood, IL: Irwin, 1949, 333 57. An often cited article establishing the Stolper Samuelson theorem, one of the basic building blocks of the Heckscher Ohlin Samuelson theory of trade.

Viner, Jacob. 1937. *Studies in the Theory of International Trade*. New York: Harper. A magisterial survey of preclassical, classical, and neoclassical trade theories.

ANDREA MANESCHI

■ gender

The global work force is largely female. First, at the international level, women make up about 50 percent of migrant workers in Asia, even 75 percent and their share is increasing. These are women from low-income countries, such as the Philippines and Sri Lanka, taking up jobs elsewhere in typically feminine occupations such as domestic work, child care, and nursing. The countries of destination are mainly Western, but also Arab countries, such as the United Arab Emirates. Second, at the national level, the majority of workers in export industries in developing countries is female. Together, these two levels of the global work force imply that more than 50 percent of workers in the global economy that is, in globalized jobs are women. This trend stands in striking contrast with the fact that universally, women's labor force participation is lower than that of men.

Female Employment in Export Manufacturing The female share of employment in export manufacturing reached its peak in the 1970s and

1980s, in particular in countries that pursued an export-led growth strategy. These countries offered export processing zones (EPZs) to (foreign) investors with incentives such as tax holidays and limited enforcement of labor regulations. The countries include Mexico, Taiwan, and South Korea, but also low-income countries such as Bangladesh, Sri Lanka, and Mauritius, that all have shown female shares of export employment between 70 and 90 percent. Since the 1990s this share has declined, but remains above 50 percent for developing countries, whereas the female share in export manufacturing in countries that belong to the Organisation for Economic Co-operation and Development (OECD) has declined and is well below 50 percent. In developed countries, export employment has declined in labor-intensive sectors, which have moved to developing countries due to the wide availability of low-skilled labor and low wages there, as is predicted by the Heckscher-Ohlin model of international trade. Women in OECD countries have particularly lost jobs in female-intensive sectors such as textiles, garments, and leather products.

The high female shares of migration and export employment show that the world economy has provided new opportunities for women in the developing world to find jobs and earn an independent income. The reasons for the predominantly female employment in export industries in developing countries relate to perceived qualities of female labor as well as to labor market discrimination.

Women tend to be regarded by employers as more suitable for work in typical low-skilled jobs in export manufacturing production. This is also referred to as the "nimble fingers" view, which leads many employers in sectors such as garments, toys, microelectronics, and assembly to prefer female over male workers. The "nimble fingers" view of women workers values the typical skills required for detailed but repetitive tasks in mass production, which have not been acquired through formal education but through women's socialization (informal training in sewing for example). In addition, the "nimble fingers" perspective of female workers also extends to the perception of women workers as being socialized as more obedient and less militant than men, which is reflected in women's lower membership in trade unions.

Wage Discrimination Women are the preferred work force in industries that compete at the international level and experience a strong pressure toward cost reduction. Women are hired more often than men in such industries, because women's wages are lower than men's wages. This gender wage gap provides employers with a cost advantage, without much loss of productivity as long as women are hired only for low-skill jobs. Gender analyses of export employment indeed show that women are predominantly found in the lowest skill categories of production, both in wage employment and in subcontracted work, such as piece-rate home work. Econometric analyses have shown that countries with the highest ratio of export earnings to their gross domestic product also have the largest gender wage gaps, even when educational levels of women have moved close to those of men. The Asian tiger economies, the fast growing economies of South Korea, Taiwan, Hong Kong and Singapore, are prominent examples, suggesting that their export success has, at least in part, been built on wage discrimination against women.

Wage discrimination is not specific to export employment, but trade liberalization does tend to put wages in export sectors under pressure, as well as wages in import competing sectors. This makes it difficult to improve women's wages and to reduce the gap with men's wages without a loss in export earnings or a worsening of the competitiveness of domestic goods vis-à-vis imported goods. Moreover, gender norms in labor markets are strong and justify women's low wages on the basis of gender stereotypes. One was mentioned above, the "nimble fingers" view of women workers, in which women's skills are highly valued for export production but not rewarded in wages because the skills are perceived as "natural" and not part of women's formal human capital. Another gender norm that prevents women's wages from catching up with men's is the breadwinner bias in employers' decisions on payment: men are thought to deserve higher wages than

women because of their traditional gender role as breadwinners. This norm, however, doesn't take into account that men's wages are often not sufficient to serve as a family wage and that globally there is an increase in the share of female-headed households, hence, an increase in the number of female breadwinners.

Women as the World's Flexible Work Force
Women are not only the preferred work force in wage employment for exports, but also make up the majority of flexible workers in the global production system. The informal economy, which is largely unregulated, with low or absent labor standards, is expanding. This is partly due to the outsourcing of production by multinational companies to local suppliers in global value chains   production processes with various layers of heavily competing suppliers for well-known brand names. Indeed, the global work force has been increasingly informalized, with women experiencing lower labor standards than men. Jobs in the global production system are increasingly subcontracted, flexible, and temporary. Women therefore find themselves increasingly working informally at the bottom of global value chains, outside the reach of labor laws, labor inspectors, and company codes of conduct.

The explanation for the increase in outsourcing rather than in-company production is, again, the cost pressure experienced by firms that try to survive in highly competitive low-wage sectors of the world economy. Large parts of production in these sectors are subcontracted and end up being done by flexible, irregular female workers in rented factories, small-scale workshops, or at home by individual workers. This process of global cost reduction through hiring cheap female labor is quite different from a process in which companies that are less vulnerable to global competition compete on the basis of productivity increases rather than cost reductions. Such a strategy requires investment in new technology and skills. Even though the educational gap between men and women is narrowing everywhere, and has disappeared in an increasing number of countries, it is men who benefit more often from the on-the-job

training and higher pay in such industries. This is exactly what is reflected in the figures referred to above, showing the defeminization of export employment in mature export industries. Employers prefer men for the higher skill jobs, believing that women would be less reliable workers because of maternity leave and possible career breaks when they have children. In addition, male workers' bargaining power is higher than that of female workers because of the male breadwinner norm and men's higher participation in trade unions. Women workers, therefore, remain at the lower end of export production.

The prospects for a reduction of the gender wage gap and improvement of working conditions are limited, despite women's increased levels of education, as long as labor market discrimination keeps women locked in low-skilled formal and informal employment in highly competitive global industries.

See also export processing zones; export promotion; fragmentation; Heckscher-Ohlin model; labor standards; migration, international; outsourcing/offshoring; social policy in open economies; trade and wages

FURTHER READING

Beneria, Lourdes. 2003. *Gender, Development, and Globa lization: Economics as If All People Mattered*. New York: Routledge. This fine political economy study about the gender dimensions of globalization provides a compre hensive overview of the main issues, including such topics as labor standards, migrant workers, and unpaid work.

Grown, Caren, Diane Elson, and Nilüfer Çağatay, eds. 2000. *World Development* 28 (7) (Special Issue on Growth, Trade, Finance, and Gender Inequality). A rich collection of short papers that integrate gender analysis in macroeconomic analysis. Not all contributions to the special issue are about the world economy, but those that are present a variety of perspectives, ranging from formal models to econometric cross country studies and policy oriented papers.

International Working Group on Gender, Macro economics, and International Economics. Website available at http://www.econ.utah.edu/genmac. Salt

Lake City: University of Utah. Features working papers, discussions, and other sources of information on gender and trade.

Seguino, Stephanie. 2000. "Gender Inequality and Economic Growth: a Cross Country Analysis." *World Development* 28 (7): 1211 30. A key publication that provides strong empirical evidence for the thesis that the success of export led growth strategy largely depends on gender wage discrimination.

Standing, Guy. 1999. "Global Feminization through Flexible Labor: A Theme Revisited." *World Development* 27 (3): 583 602. An update of an earlier paper, showing that women make up the large part of the global work force, but under worse conditions and lower wages compared to men.

Staveren, Irene van, Diane Elson, Caren Grown, and Nilüfer Çağatay, eds. 2007. *The Feminist Economics of Trade*. London: Routledge. The first volume that brings together research on trade and gender. It analyzes not only gender impacts of trade but also shows how trade is affected by preexisting gender inequalities.

IRENE VAN STAVEREN

General Agreement on Tariffs and Trade (GATT)

From its creation in 1947 to its absorption into the newly created World Trade Organization (WTO) in 1995, the General Agreement on Tariffs and Trade universally referred to as "the GATT" was *the* multilateral agreement governing trade among a growing number of countries, including all of the major trading nations. As such, the GATT served two primary functions for its members. First, it provided a set of multilaterally agreed rules and disciplines governing selected policies affecting countries' trade in goods, in particular, tariffs and quotas on imports. The goal was to encourage international trade by imposing certain disciplines on countries' trade policies, thereby making future access to foreign markets more predictable. Second, it provided a forum for (1) the settlement of trade disputes among members, and (2) negotiations to further liberalize

trade, as well as to strengthen and extend the multilateral rules. Thus the GATT was both an agreement and an institution.

By any standard, the period 1948 to 1994 was an era of unparalleled economic prosperity for the world economy. Adjusted for inflation, output of goods and services grew at an average annual rate of 3.6 percent, with output per person increasing 2 percent a year (compared with an estimated 0.9 percent during 1820 1913). A key factor behind this rapid growth of output was the even more rapid expansion of world merchandise trade, which grew at an average rate of nearly 6 percent a year in real terms from 1948 to 1994. This meant that virtually every year saw a growing share of world production being traded across one or more borders.

It is widely accepted among trade specialists that GATT rules, in combination with periodic GATT-sponsored negotiations, played a major role in this expansion of world trade (Goldstein, Rivers, and Tomz 2006). The GATT has also attracted considerable interest from scholars and policy officials who believe a good case can be made that the GATT was the most successful of the postwar international economic institutions.

Origins of the GATT The immediate origin of the GATT was the UN draft charter for the stillborn International Trade Organization, often referred to as the "Havana Charter." Before going into this, however, it is useful to mention two earlier experiences with international commercial policies that influenced the drafters of the Havana Charter and thus the ultimate design of the GATT.

In the latter half of the 19th century, European treaties increasingly included a provision for most-favored-nation (MFN) trade status. Enshrined in trade agreements such as the Cobden-Chevalier Treaty of 1860 liberalizing trade between Great Britain and France, the MFN provision assured that trade was nondiscriminatory. This meant that trade benefits resulting from subsequent treaties were passed on to previous treaty partners through the promise that they would receive the lowest tariff rate granted to any trade partner. MFN created a

multiplier effect for trade liberalization, leading to significantly lower barriers to trade throughout the continent (Curzon 1965).

More important, the drafters of the Havana Charter shared the experience of the Great Depression, when economic decline brought beggar-thy-neighbor policies, high tariffs, and depreciating currencies, as nations tried to protect their home markets. This, of course, only accentuated the general decline. The extent of the worldwide descent into virtually uncontrolled protectionism is evident from the 25 percent decline in the volume of world trade between 1929 and 1932 (minus 60 percent in dollar terms). By 1936 industrial production had recovered to a level 10 percent *above* its 1929 level, while the volume of world trade was still 15 percent *below* its 1929 level.

It was against this background that, in the closing years of World War II, American and British officials began laying plans for the postwar world economy. Among the most ambitious of those plans was the proposal for a new institution, to be called the International Trade Organization (ITO), that would be the third pillar along with the International Bank for Reconstruction and Development (World Bank) and the International Monetary Fund (IMF), both already agreed to in 1944 underlying the postwar world economy. A preparatory committee established by the UN Economic and Social Council met in London in October November 1946 to discuss a draft charter for the proposed ITO. That meeting also produced a resolution calling for a multilateral trade negotiation, to be held the following year in Geneva at the committee's second session and a recommendation that a "General Agreement on Tariffs and Trade," containing rules and disciplines to protect the value of the tariff concessions, be created.

The first full draft of the GATT, based primarily on text taken from the commercial policy section of the draft ITO charter, was produced at a January February 1947 meeting in Lake Success, New York (Jackson 1969). Two concurrent negotiations began in Geneva in April, one focused on the draft ITO charter and another focused on reductions in tariffs and refinements in the accompanying draft GATT.

On October 30, 1947, 13 developed and 10 developing countries accounting for more than three-quarters of world trade signed the Final Act of the GATT containing (1) a schedule of concessions listing each government's agreed tariff reductions, and (2) rules and procedures regulating nontariff restrictions on trade what Robert Hudec (1987) called a "code of behavior" for governments.

It was assumed that the ITO, whose charter was finalized in Havana in March 1948, would shortly come into force and that the GATT would come under its aegis. That option evaporated in 1950 when it became clear that the U.S. Congress would not ratify the Havana Charter. The provisional GATT became the framework for postwar commercial relations. Noting that the early GATT "was permeated by an atmosphere of impermanence," Richard Gardner (1956) described it as "a slender reed on which to base progress toward a multilateral regime."

The 23 participants in the Geneva trade negotiations were the GATT's first "contracting parties" a term that emphasized both the contractual nature of the GATT and the fact that it was a "provisional agreement" and not a treaty-based organization. With a more or less steady increase in the membership over the period 1948 94, the number of contracting parties stood at 128 when the GATT was absorbed into the WTO in 1995, four-fifths of which were developing or least-developed countries.

Basic Principles Five basic principles some explicit, some implicit informed the rules and guided the activities of the contractual, rules-based GATT system.

Nondiscrimination. This principle embraces two elements. The better-known one is the most-favored-nation (MFN) clause, which prohibits a GATT member from discriminating between imports from other GATT members any concession granted on a product from one contracting party "shall be accorded immediately *and unconditionally* to the like product originating in or destined for the territories of all other contracting parties" (Article I, emphasis added). Adherence to MFN both depoliticizes trade and ensures that a country buys its imports from the cheapest foreign source. Some exceptions were in-

cluded in the GATT from the beginning, most notably for free trade areas and customs unions among small groups of countries. Later the contracting parties agreed to allow preferential treatment of imports from developing countries. By the early 1990s, with the rapid spread of preferential trading areas especially free trade areas many trade experts were openly questioning the relevance of the GATT's MFN provisions.

The second element of nondiscrimination is national treatment, which prohibits discrimination between imports and domestic products once the imports have entered the country. For example, a retail sales tax applied to imported TV sets but not to domestically produced TV sets would violate the national treatment principle (however, imposing a tariff on foreign-made TV sets as they enter the country is not a violation of national treatment).

Tariffs only. The GATT was not a free trade organization. Member countries were not required to participate in trade liberalizing negotiating rounds, and they were free to protect domestic producers by imposing restrictions on imported goods. If they did protect, however, they were required to use tariffs rather than import quotas or other nontariff restrictions to provide that protection. This made good sense. Not only is the level of protection provided by tariffs more transparent than nontariff protection, but tariffs are a less distorting (less inefficient) way of granting protection.

Consensus decision making. Although there were provisions for majority voting on a one-country, one-vote basis, the GATT developed a very strong tradition of making all decisions by consensus (defined as no country present at the meeting when the decision was taken formally objecting to the proposed decision). This often slowed the decision-making process, but considering that many of the decisions resulted in legally binding obligations for all members had important advantages, including adding to the intellectual and diplomatic "legitimacy" of the decisions.

Special treatment for developing countries. There were no special provisions for developing countries in the original GATT. In the 1950s the developing country members began demanding exemptions from GATT's rules and disciplines, to which was added, at the beginning of the 1960s, demands for special and more favorable treatment for their exports in the markets of other GATT members (Hudec 1987). Institutional changes reflecting these demands included special provisions for developing countries in GATT rules and procedures, the creation of a permanent Committee on Trade and Development, the Tokyo Round's "Framework Agreements," and technical assistance provided by the GATT Secretariat. Technical assistance aside, the extent to which these various developments actually benefited rather than harmed developing countries continues to be the subject of debate among trade experts.

Small Secretariat with a limited mandate. The member countries kept the Secretariat very small compared with the other major postwar international economic organizations (such as the World Bank and the IMF). Comparable figures for 1994 are not available, but in 1996, just after GATT's absorption into the WTO, the WTO Secretariat was in 16th place on the list of the 17 largest international economic organizations in terms of budget and number of staff members (Blackhurst 1998).

As for the Secretariat's limited mandate, consider, for example, the following: the Director-General could not initiate a dispute settlement case, no matter how blatant the rule violation; the Secretariat could not interpret GATT rules; the Secretariat was not allowed to report regularly on trade policy developments in member countries until 1977, 30 years after the GATT was signed; and the Secretariat was not allowed to have a formal Office of Legal Affairs until 1983, even though the GATT was based on contractual obligations. In essence, the Secretariat existed to service the meetings and other activities of the member countries. The popular way to describe this situation was to say that the GATT was a "member-driven" institution.

The explanation for the member-driven operation of the GATT and the Secretariat's very small size undoubtedly is the highly politicized nature of trade policies at the national level. Domestic politics meant

that no government and especially no developed country government could be seen advocating a secretariat with any real independence and power. Some analysts argue that this was the secret behind GATT's success (Hudec 1993; Winham 1998).

Dispute Settlement A rules-based agreement is credible only if it includes effective dispute settlement procedures. In GATT's early years, working parties were created to handle disputes. Then in 1952 the contracting parties switched over to a "panel" system in which three (sometimes five) delegates from countries not party to the dispute evaluated submissions by the parties to the dispute and made recommendations or rulings, depending on the issue(s) at stake. The panel's final report was submitted to the GATT Council (a body representing all GATT members) for adoption by consensus, which meant that the country that "lost the case" could block the adoption of the panel report if it so desired. Odd as this may sound, the system worked in the first three decades a substantial proportion of panel reports was adopted because the contracting parties knew that if the dispute settlement system failed, the GATT would fail (blockage of panel reports emerged as a major problem only in the 1980s and was addressed in the Uruguay Round negotiations).

A defendant country that lost a case (and did not block the adoption of the panel's report) had three options: (1) withdrawal of the GATT-inconsistent measure(s), always the required and first-best option; (2) as an interim measure, pay compensation, generally by lowering one or more tariffs (on an MFN basis) on selected exports of interest to the winning plaintiff; or (3) if compensation was not paid, suffer retaliation by the plaintiff country, generally via a discriminatory increase in tariffs on items of export interest to the defendant country. The third option was possible only if it was authorized by the contracting parties a very important proviso since it heavily modified the long-standing "right of unilateral reprisals" in international law. Although retaliation was authorized only once under the GATT, there is no doubt that countries' awareness of this option was crucially important in the enforcement of GATT rules.

In the early years there was a struggle between those countries that favored an approach to dispute settlement that was pragmatic and avoided undue "legalism" and those that favored a more judicial or legalistic approach. Over time the latter approach won out, as evidenced, for example, by the previously mentioned creation of GATT's first ever Office of Legal Affairs in 1983. In the course of providing other examples, Ernst-Ulrich Petersmann (1997) observes that this shift progressively transformed "the GATT into the most frequently used, and most effective, international system for the rule-oriented settlement of disputes among governments."

Beginning in the 1980s, the effectiveness and uniqueness of GATT's dispute settlement system acted as a magnet for issues outside GATT's area of responsibility. In two instances the protection of intellectual property and the enforcement of workers' rights proposals to extend GATT rules and dispute settlement (with trade retaliation a possibility) into new areas were motivated by the lack of effective enforcement mechanisms in the responsible international organizations, while in two others trade in services (banking, insurance, and so forth) and protection of the environment there was no effective international organization. Negotiations on these issues subsequently became key parts of the Uruguay Round (see below).

Negotiating Rounds Although the GATT was not a free trade organization, the underlying "ethos" included a strong belief by many members including all the developed countries that the pursuit of *freer* trade was an important goal for the organization. This is not surprising considering the high tariffs and widespread use of import quotas, inherited from the 1930s, that confronted the early GATT.

During the period 1947 94, GATT's contracting parties held eight rounds of trade liberalizing negotiations. Reciprocal reductions in tariffs I'll reduce my tariff on radios if you reduce your tariff on corn were the exclusive focus of the first five rounds and an important part of the agendas of the three later rounds. By 1994 the average tariff on manufactured goods entering the developed countries had been reduced from around 40 percent in the immediate

postwar period to below 5 percent. Tariffs in many developing countries also started to come down beginning in the late 1960s and early 1970s as these countries adopted more outward-oriented development strategies.

Not only were tariffs reduced, but an increasing number were also "bound" under GATT rules. Binding a particular tariff, say the tariff on radios at 20 percent, meant that you were not permitted to increase it above that level (a lower tariff was okay). Raising it above 20 percent would put you in violation of your GATT obligations and risk a dispute settlement case. Binding members' tariffs was often as important as reducing rates since both increased the predictability of future market access and thus encouraged trade.

In the Kennedy Round (1964 67) anti-dumping measures were added to the agenda. Reflecting both the success in reducing tariffs and the increasing complexity of commercial relations, the Tokyo Round (1973 79) attempted, for the first time, to deal with a range of nontariff measures, including three production subsidies, technical barriers to trade (for example, safety standards), and government procurement that are applied not at the border but domestically in the importing country. The extent to which these initial steps "inside the border" signaled a turning point for the multilateral trading system was perhaps not fully appreciated at the time. Otherwise, little meaningful progress was made in the Tokyo Round, the principal achievement being a series of "codes" that were signed by a minority of the contracting parties, making them plurilateral rather than multilateral agreements.

The Uruguay Round It was not long before pressure to launch a new round began building as a result of both short-term concerns and medium-term goals. The former focused on the worsening trade relations among many of the leading traders in the early 1980s, caused in part by a sharp slowdown in the growth of world production and trade. Launching a new round, it was argued, would help governments to contain protectionist pressures, through the new round's standstill agreement (on new trade barriers) and the demonstration that governments were dealing with the system's perceived shortcomings.

The medium-term goals behind the push for a new round, in turn, fell into two distinct categories. One was based on the widespread perception that the GATT system was in serious trouble and that something had to be done soon. Not only had the postwar liberalization momentum been lost, but even more ominously, the use of GATT-illegal discriminatory quantitative restrictions (often labeled voluntary export restraints, or VERs) by some of the major countries was spreading to ever more products, as was the use of trade-distorting production subsidies. Moreover, after seven rounds of negotiations, agricultural trade was still not subject to GATT's rules, and trade in textiles and clothing a potentially important source of jobs and economic growth in many lower-income developing countries remained heavily restricted by GATT-sanctioned quantitative restrictions.

The other medium-term goal stemmed from the belief on the part of many members especially the developed countries that after nearly 40 years the GATT needed to be updated to be able to deal effectively with the new realities of world trade and the policy frictions they engendered. A key part of this updating involved proposals for writing new GATT rules and disciplines to cover economic activities that had increased greatly in importance since the GATT was drafted. Attention focused on three areas: trade in services, the protection of intellectual property rights, and trade-related investment measures (for example, requiring foreign-owned firms to purchase a certain percentage of their inputs in the host country).

With growing support from the developed countries and many developing countries, momentum for launching a new round began accelerating in the early months of 1985. At the same time, a group of developing countries led by Brazil and India remained strongly opposed to a new round. Proposals to include "new areas" on the negotiating agenda especially trade in services also drew strong opposition from many developing countries. Eventually the countries favoring a new round won out, helped

by a compromise whereby the services negotiations would be on the agenda but kept completely separate from the negotiations on goods.

GATT's eighth and final round of negotiations was launched in September 1986 at a meeting of the contracting parties in Punta del Este, Uruguay. The ministerial declaration launching the Uruguay Round gave the negotiators four years to complete their work. That target was missed when the December 1990 ministerial meeting in Brussels failed to resolve differences that had blocked agreement on a number of difficult issues, of which agricultural trade was by far the most high profile and politically charged. Negotiations continued and, finally, late in November 1993, GATT's new Director-General, Peter Sutherland, decided that there had been enough progress for him to give the negotiators a deadline of December 15 to wind up the round. The strategy worked. Describing the meeting of the Trade Negotiations Committee on December 15, John Croome (1995) observes, "At 7:30 in the evening, to huge applause, Sutherland brought down his gavel to signify approval of the Uruguay Round agreements. . . . It was, he suggested, 'a defining moment in modern economic and political history.'"

On a very general level, the principal achievements of the Uruguay Round can be summarized as follows:

- Agreement to create the World Trade Organization to replace the 1947 GATT provisional agreement.
- A major revision of the dispute settlement system. In a complete reversal of the practice under the GATT, panel decisions are now automatically accepted unless there is a consensus not to do so; this was coupled with the creation of an independent Appellate Body to hear appeals of panel decisions.
- Further liberalization of tariffs and nontariff restrictions on industrial products, and—for the first time—reductions in import barriers and other (domestic) interventions affecting agricultural trade.
- Creation of the General Agreement on Trade in Services (GATS)
- Creation of the Agreement on Trade-Related Aspects of Intellectual Property Rights (the TRIPS Agreement)

Trade ministers met in Marrakesh (Morocco) in April 1994 to sign the Uruguay Round Agreements, and the WTO began its existence on January 1, 1995.

What Became of the GATT? The original agreement, now referred to as GATT 1947, was phased out at the end of 1995. Within the WTO family of agreements, an updated version of the agreement GATT 1994 is the umbrella treaty covering trade in goods.

The GATT played a key role in the economic prosperity of the second half of the 20th century. It also gave us the WTO, modeled on the GATT but with greatly expanded responsibilities in an increasingly globalized world economy. On both counts, we owe a great debt to the architects of the GATT and to the many national officials and Secretariat staff who despite a very inauspicious beginning made it work.

See also Agreement on Agriculture; Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS); agricultural trade negotiations; anti-dumping; General Agreement on Trade in Services (GATS); labor standards; multilateralism; nondiscrimination; nontariff measures; quotas; special and differential treatment; tariffs; technical barriers to trade; textiles and clothing; trade in services; Uruguay Round; voluntary export restraints; World Trade Organization; World Trade Organization dispute settlement

FURTHER READING

Barton, John, Judith Goldstein, Timothy Josling, and Richard Steinberg. 2006. *The Evolution of the Trade Regime: Politics, Law, and Economics of the GATT and the WTO.* Princeton: Princeton University Press. An excellent interdisciplinary overview of the GATT regime and its transformation into the WTO.

Blackhurst, Richard. 1998. "The Capacity of the WTO to Fulfill Its Mandate." In *The WTO as an International Organization*, edited by Anne O. Krueger. Chicago: University of Chicago Press, 31 58. Although con

cerned mainly with the WTO, this paper by the former Director of Economic Research at both the GATT and WTO Secretariats (1985 97) includes a relatively de tailed inside view of the operation of the GATT Secre tariat.

Croome, John. 1995. *Reshaping the World Trading System: A History of the Uruguay Round.* Geneva: WTO Secre tariat. An engaging, step by step, firsthand account by a former senior WTO official of the preparatory phase and the seven years of complex and wide ranging negotiations.

Curzon, Gerard. 1965. *Multilateral Commercial Diplomacy.* Middlesex, UK: Michael Joseph. This early book on the GATT by an academic economist offers a careful anal ysis of how the GATT evolved from its inauspicious beginning as a strictly interim arrangement into an ef fective multilateral institution.

Gardner, Richard N. 1956. *Sterling Dollar Diplomacy.* Oxford: Oxford University Press. A standard reference work on both the steps leading to the creation of the GATT and its activities in the early years.

General Agreement on Tariffs and Trade. 1985. *Trade Pol icies for a Better Future.* Geneva: GATT Secretariat. A report by seven independent world figures containing 15 recommendations for dealing with a variety of challenges facing the GATT in the first half of the 1980s. The report is generally credited with playing an important role in preparations for the Uruguay Round.

Goldstein, Judith, Douglas Rivers, and Michael Tomz. 2006. "Institutions in International Relations: Under standing the Effects of GATT and the WTO on World Trade.*" International Organization* 61: 2. This article assesses the trade implications of membership in the GATT and provides empirical evidence that GATT membership increased trade for both developed and developing countries.

Goode, Walter. 2003. *Dictionary of Trade Policy Terms.* 4th ed. Cambridge: Cambridge University Press, World Trade Organization, and the Center for International Economic Studies (University of Adelaide). A nearly indispensable guide to the jargon filled world of the GATT, the WTO, and other international economic organizations. The volume also covers various GATT/ WTO rules, newer issues, and developing country con cerns.

Horn, Henrik, and Petros C. Mavroidis, 2001. "Economic and Legal Aspects of the Most Favored Nation Clause." *European Journal of Political Economy* 17 (2): 233 79. This article examines the merits of the theoretical ra tionales for MFN and tests them against the background of the current regulatory framework.

Hudec, Robert E. 1987. *Developing Countries in the GATT Legal System.* Aldershot, UK: Gower for the Trade Policy Research Centre (London). An excellent history, by a leading GATT legal scholar, of the GATT's evolving legal relationship with its developing country members, coupled with a critique of the relationship as it stood prior to the GATT's absorption into the WTO.

———. 1993. *Enforcing International Trade Law: the Evolution of the Modern GATT Legal System.* Salem, NH: Butterworth Legal Publishers. A standard reference work, by a leading GATT legal scholar, on the GATT legal system as it stood just prior to its major overhaul in 1995 as part of the Uruguay Round Agreements.

Irwin, Douglas A. 2005. *Free Trade under Fire.* 2d ed. Princeton: Princeton University Press. An excellent re view of the case for free (freer) trade that also examines the various arguments for protection. Chapter 2 contains a nontechnical summary of the various ways trade lib eralization stimulates economic growth.

Jackson, John H. 1969. *World Trade and the Law of the GATT.* Indianapolis, IN: Bobbs Merrill. A standard reference work, by a leading GATT legal scholar, de tailing the evolution of the GATT legal system during its first two decades.

League of Nations, Economic, Financial and Transit De partment. 1942. *Commercial Policy in the Interwar Period: International Proposals and National Policies.* Geneva: League of Nations. A firsthand account of developments during 1918 39, with a focus on the many conferences and other international activities that were unable to prevent the catastrophic slide into worldwide protec tionism and economic depression.

Petersmann, Ernst Ulrich. 1997. *The GATT/WTO Dispute Settlement System.* London: Kluwer Law International. A very useful book by a leading GATT/WTO scholar that not only analyzes the GATT dispute settlement system (as well as that of the WTO), but also (a) places it in the broader context of international law and inter national organizations in general, and (b) uses several

concrete dispute settlement cases to illustrate GATT/WTO law.

Winham, Gilbert R. 1998. "The World Trade Organization: Institution Building in the Multilateral Trade System." *World Economy* 21 (3): 349 68. A political scientist's view of the operation of the GATT and its implications for the WTO.

World Trade Organization. 1999. *Guide to the Uruguay Round Agreements*. Geneva: WTO Secretariat. In terms of rhetoric and organization the official legal texts of the Uruguay Round results are virtually incomprehensible to the nonexpert. This book by WTO Secretariat staff members presents the results in an authoritative but understandable way that also makes it easy to get a good sense of the overall package.

. Website, http://www.wto.org. A comprehensive user friendly web site which, although focused on the WTO, also includes material dealing with the GATT, such as GATT documents and a list of the 128 GATT members at the time of its absorption by the WTO.

RICHARD BLACKHURST

■ General Agreement on Trade in Services (GATS)

The General Agreement on Trade in Services (GATS), an outcome of the Uruguay Round of trade negotiations, contains the first set of multilateral rules for trade in services. It entered into force on January 1, 1995, nearly fifteen years after it was first put on the multilateral trade agenda at the initiative of the United States (Feketekuty 1988). Developing countries initially resisted negotiating services, but today some, such as India, that have a growing stake in services exports, are the main proponents. After nearly twenty-five years of virtually continuous negotiations, the GATS has helped create a more secure, rule-based environment for services trade, but it has so far failed to generate significant new trade liberalization.

The Scope of the GATS The GATS is extremely wide in scope and covers all measures affecting *trade in services*. Instead of worrying about a precise definition of what a service is, GATS negotiators proceeded to list the entire range of services covered:

1. Business services
2. Communication services
3. Construction services
4. Distribution services
5. Educational services
6. Environmental services
7. Financial services
8. Health-related and social services
9. Tourism and travel-related services
10. Recreational, cultural, and sporting services
11. Transport services

As if this list were not long enough, there is a twelfth residual category, "Other services not elsewhere included."

Since the conventional definition of trade where a product crosses the frontier would miss out on a whole range of international transactions, the GATS also takes an unusually broad view of trade, which is defined (in Article I) to include four modes of supply:

- *Mode 1—Cross-border*: services supplied from the territory of one member into the territory of another. An example is software services supplied by a supplier in one country through mail or electronic means to consumers in another country.
- *Mode 2—Consumption abroad*: services supplied in the territory of one member to the consumers of another. An example is when the consumer moves, for instance, and consumes tourism or education services in another country. Also covered are activities such as ship repair abroad, where only the property of the consumer moves.
- *Mode 3—Commercial presence*: services supplied through any type of business or professional establishment of one member in the territory of another. An example is an insurance company owned by citizens of one country establishing a branch in another country.
- *Mode 4—Presence of natural persons*: services supplied by nationals of one member in the

territory of another. This mode includes both independent service suppliers, and employees of the services supplier of another member. Examples are a doctor of one country supplying through his physical presence services in another country, or the foreign employees of a foreign bank. Note that the GATS does not apply to measures affecting natural persons seeking access to the employment market of a member, or to measures regarding citizenship, residence, or employment on a permanent basis.

Thus any measure affecting the supply of services through any of these modes is covered by the GATS. The inclusion of commercial presence as a mode extends the reach of the agreement to measures affecting foreign direct investment (FDI), and the inclusion of presence of natural persons (at the insistence of developing countries) to measures affecting the entry of foreign nationals, both of which have traditionally been a tightly controlled province of national governments.

The Rules of the GATS The major provisions of the GATS are summarized in table 1. The wide scope of the GATS contrasts with the gentleness of its rules. It is convenient to think of GATS rules as operating at two levels. First, there is a set of general rules that apply across the board to measures affecting trade in services; and second, there are the sector-specific commitments made by members, which determine the liberalizing impact of the agreement.

Key General Rules: The MFN Principle and Transparency Article II of the GATS constitutes a general obligation which is, in principle, applicable across the board by all members to all services sectors. Article II:1 of GATS states: "With respect to any measure covered by this Agreement, each Member shall accord immediately and unconditionally to services and service suppliers of any other Member treatment no less favorable than that it accords to like services and service suppliers of any other country." The GATS and its Most Favored Nation (MFN) obligation came into effect, however, before World Trade Organization (WTO) members were willing to completely eliminate discriminatory measures in services trade.

Table 1
Major provisions of the GATS
(article and main disciplines implied)

I	Definition. Trade in services covers all four modes of supply.
II	Most Favored Nation (MFN) obligation. Option to invoke exemptions on a one time basis.
III	Notification and publication. Obligation to create an enquiry point.
IV	Increasing participation of developing countries. High income countries to take measures to facilitate trade of developing nations.
V	Economic integration. Allows for free trade and similar agreements.
VI	Rules for domestic regulation. Requirements concerning the design and implementation of service sector regulation, including in particular qualification and licensing requirements.
VII	Rules on recognition of qualifications, standards, and certification of suppliers.
VIII	Monopolies and exclusive suppliers. Requires that such entities abide by MFN and specific commitments (Articles XVI and XVII) and do not abuse their dominant position.
IX	Business practices. Recognition that business practices may restrict trade. Call for consultations between members on request.
XIV	General exceptions. Allows measures to achieve noneconomic objectives.
XVI	Market access. Defines a set of policies that may only be used to restrict market access for a scheduled sector if they are listed in a member's specific commitments.
XVII	National treatment. Applies in a sector if a commitment to that effect is made and no limitations or exceptions are listed in a member's schedule.
XVIII	Additional commitments. Allows for any other specific commitment to be made on a sector by sector basis. To date these have been limited primarily to telecommunications, through the so called Reference Paper.
XIX	Calls for successive negotiations to expand coverage of specific commitments (Articles XVI and XVII).
XXIX	States that annexes are an integral part of the GATS.

Specific sectoral sensitivities, such as a large number of bilateral agreements in maritime and road transport, that were revealed in the Uruguay Round raised the specter of wholesale sectoral exclusions from GATS as a means of avoiding the MFN rule. In order to prevent this, it was agreed to permit limited exemptions to MFN under GATS. Such exemptions, however, had to be taken at the time the negotiations were concluded and *in principle* were not meant to last longer than 10 years (that is, not beyond 2004) but at present they show no sign of disappearing.

Apart from services specified in individual MFN exemption lists, the only permitted departure from most-favored-nation treatment under the GATS covers preferential treatment among countries that are members of regional trading arrangements. The GATS rules on "Economic Integration," in Article V, are modeled on those in Article XXIV of the GATT. Article V:1 permits any WTO member to enter into an agreement to further liberalize trade in services with the other countries that are parties to the agreement, provided the agreement has "substantial sectoral coverage," eliminates measures that discriminate against service suppliers of other countries in the group, and prohibits new or more discriminatory measures. An approved agreement must be designed to help trade among its members, and must not result in an increase in the overall barriers faced by nonmembers in trading with the group within the respective sectors or subsectors (Article V:4). If the establishment of the agreement, or its subsequent enlargement, leads to the withdrawal of commitments made to nonmembers, there must be negotiations to provide appropriate compensation (Article V:5). Agreements that have been notified so far include the North American Free Trade Agreement; the European Communities; agreements between the European Union (EU) and the Slovak Republic, Hungary, Poland, the Czech Republic, Romania, Norway, Iceland, Liechtenstein, and Bulgaria (many of which became redundant upon accession of the countries concerned to the EU); and agreements between Canada and Chile and between Australia and New Zealand.

A related exception from the MFN rule, for the movement of natural persons, is permitted by Article V *bis* of the GATS. This allows countries to take part in agreements that establish full integration of labor markets. The only such agreement notified so far is the one involving Denmark, Finland, Iceland, Norway, and Sweden.

Specific Commitments on Market Access and National Treatment The GATS depends to a large extent on the specific commitments on market access and national treatment made by members. Both types of commitments are made for each of the four modes of delivery of service transactions.

Article XVI stipulates that measures restrictive of *market access* that a WTO member cannot maintain or adopt, unless specified in its schedule, include limitations on: (1) the number of service suppliers; (2) the total value of services transactions or assets; (3) the total number of services operations or the total quantity of service output; (4) the total number of natural persons that may be employed in a particular sector; (5) specific types of legal entity through which a service can be supplied; and (6) foreign equity participation (e.g., maximum equity participation). With the exception of (5), the measures covered by Article XVI all take the form of quantitative restrictions.

Three aspects of Article XVI are important. First, the Article XVI list does not include all measures that could restrict market access. Perhaps most significantly, fiscal measures are not covered. Thus, a member could maintain, without being obliged to schedule, a high nondiscriminatory tax on a particular service that severely limits market access. Second, Article XVI has been interpreted to cover both discriminatory and nondiscriminatory measures, that is, measures of the type "only five new *foreign* banks will be granted licenses" and also measures such as "only ten new [*foreign and domestic*] banks will be granted licenses." Finally, the limitations must be read as "minimum guarantees" rather than "maximum quotas," in other words, a country that has promised to allow five foreign banks entry is free to grant entry to more than five.

The other key pillar of the GATS is the *national treatment* obligation. Article XVII:1 states: "In the sectors inscribed in its Schedule, and subject to any conditions and qualifications set out therein, each

Member shall accord to services and service suppliers of any other Member, in respect of all measures affecting the supply of services, treatment no less favorable than that it accords to its own like services and service suppliers." Unlike Article XVI, Article XVII provides no exhaustive list of measures inconsistent with national treatment. Nevertheless, Article XVII:2 makes it clear that limitations on national treatment cover cases of both de jure and de facto discrimination. If domestic suppliers of audiovisual services are given preference in the allocation of frequencies for transmission within the national territory, such a measure discriminates explicitly on the basis of origin of the service supplier and thus constitutes formal, or de jure, denial of national treatment. Alternatively, a measure stipulating that prior residency is required to obtain a license to supply a service does not formally distinguish service suppliers on the basis of national origin, but it may de facto offer less favorable treatment for foreign suppliers because they are less likely to be able to meet a prior residency requirement than like service suppliers of national origin.

A member's specific commitments can be seen as the outcome of a two-step decision. Each member first decides which service sectors will be subject to the GATS market access and national treatment disciplines. It then decides which measures violating market access and/or national treatment respectively will be kept in place for each mode in that sector. Granting unrestricted market access with full national treatment would be equivalent to establishing free trade, and the flexible structure of rules reflects the desire of most governments to adopt a gradual and conditioned approach to opening up their markets. One result of this structure is that the GATS is not a particularly transparent or user-friendly instrument. More important, virtually all commitments made in the Uruguay Round were at best of a standstill nature, that is, a promise not to become more restrictive than already was the case for scheduled sectors (Hoekman 1996).

Doha Agenda and Beyond The negotiations under the Doha agenda show little sign of advancing services liberalization in the manner anticipated in the Uruguay Round. The best current offers of access in services do not even reflect the liberalization that has already taken place. The "request-offer" negotiating process, conducted first bilaterally and then plurilaterally, seems to have resulted in a low-level equilibrium trap where little is expected and less offered.

The WTO's Hong Kong Ministerial in December 2005 did identify the elements of a meaningful outcome in services. Such a package would have three elements. First was a promise not to impose new restrictions on trade in services. This would dispel the specter of protectionism that hangs over outsourcing of business services   which is producing huge cost savings in the North and ever-widening export opportunities for the South. Second was a commitment to eliminate barriers to FDI, either immediately or, in sectors where regulatory inadequacies need to be remedied, in a phased manner. The greatest benefits of securing openness to FDI, especially in infrastructure services, would accrue to the South while offering increased business opportunities to the North. Third was agreement to allow somewhat greater freedom of international movement for individual service providers (mode 4) in order to fulfill specific services contracts. Research, for example, by Winters et al. (2003) shows large potential benefits to both the North and the South from the liberalization of mode 4, as it offers a way to realize the gains from trade while averting social and political costs in host countries and brain drain losses for source countries.

For there to be a reasonable prospect of achieving these goals, more attention needs to be given to the regulatory context in which services liberalization takes place, in three ways. First, it may make sense to focus primarily on securing "national treatment," that is, ending all discrimination on the entry and operation of foreign services providers rather than on creating more intrusive disciplines. This will reassure regulators that multilateral commitments deprive them only of the freedom to discriminate and do not limit their freedom to regulate in any other way or to adopt policies that improve sector performance.

Second, it may be necessary to establish a credible mechanism to provide regulatory assistance to support liberalization commitments by developing

countries. This will reassure developing country policymakers that regulatory inadequacies that could undermine the benefits of liberalization will be remedied before any market-opening commitments take effect.

Third, it should be possible to make temporary entry of foreign services providers conditional on the fulfillment of specific conditions by source countries. Immigration authorities in host economies must be assured that source countries will cooperate to screen services providers, to accept and facilitate their return, and to combat illegal migration.

In essence, these proposals suggest that it may be necessary for services negotiations to be complemented by broader cooperation. The assumption is that poorer developing countries would participate meaningfully in negotiations that offered an opportunity not merely to make binding commitments but also to mobilize assistance for regulatory reform. And migration authorities would engage constructively in negotiations that offered the opportunity not only to untie their hands, but also to secure assistance from source countries to deal with problems they cannot solve on their own.

See also Doha Round; financial services; nondiscrimination; temporary movement of natural persons; trade in services; World Trade Organization

FURTHER READING

Feketekuty, Geza. 1988. *International Trade in Services: An Overview and Blueprint for Negotiations*. Cambridge, MA: Ballinger. One of the first proposals for services negotiations at the WTO by one of the main protagonists.

Hoekman, Bernard. 1996. "Assessing the General Agreement on Trade in Services." In *The Uruguay Round and the Developing Countries*, edited by Will Martin and L. Alan Winters. Cambridge: Cambridge University Press, 88–124. The first and most comprehensive assessment of the GATS agreement and commitments.

Mattoo, Aaditya. 2005. "Services in a Development Round: Three Goals and Three Proposals." *Journal of World Trade* 39 (6): 1223–38. Proposals on how the GATS could become a more powerful catalyst for services trade reform.

Winters, L. A., T. L. Walmsley, Z. K. Wang, R. Grynberg. 2003. "Liberalizing Temporary Movement of Natural Persons: An Agenda for the Development Round." *World Economy* 26 (8): 1137–61. Estimates the potential gains from the OECD countries allowing increased immigration.

AADITYA MATTOO

■ **generalized system of preferences (GSP)**
See special and differential treatment

■ **Gini coefficient**
See global income inequality

■ **Global Environment Facility**

The Global Environment Facility (GEF) grew out of an awareness in the 1980s of transboundary environmental problems and the accompanying recognition that efforts to improve matters would be costly (Sjöberg 1994). Launched in 1991 by the World Bank as an experimental facility, the GEF entered a three-year pilot program to test approaches and resolve competing governance schemes. Today it is the single largest grant-making institution for global environmental programs (Clémençon 2006). The GEF has allocated some $5 billion for more than 1,500 projects in 140 countries (GEF 2005). Because of its size, the GEF has great potential for affecting international environmental health, but as of 2007 its clearest results had been short-term, quantifiable projects such as those within the ozone program. The GEF's ability to work with populations around the world to create integrative, sustainable, long-term benefits will bolster its own position as an international environmental lender and promote the health of environments and communities worldwide.

Mission, Funding, and Change GEF projects are developed and financed through three implementing agencies: the World Bank, UN Environment Program (UNEP), and UN Development Program (UNDP). The World Bank is responsible for investments and mobilizing resources from the private

sector. UNEP supports developing scientific and technical analysis, advancing environmental management in GEF financed activities, and managing the Scientific and Technical Advisory Panel. UNDP is responsible for building the human and institutional capacities that enable governmental agencies or nongovernmental organizations to take the actions necessary for global environmental protection. Executing agencies include regional development banks and a number of specialized UN agencies. GEF operations are coordinated by a Secretariat in Washington, DC.

The GEF provides concessional financing to cover the incremental costs necessary to achieve global environmental benefits. Incremental costs are calculated by subtracting the costs of any national or local benefit from the total cost of the project in order to provide financial incentives to create global environmental benefits. Cofunding is expected to cover the "national" benefits of the project. As such, GEF funding is intended to complement traditional development assistance by covering the additional costs or agreed incremental costs incurred when a development project also targets global environmental objectives. As one observer noted, "GEF assistance is not development 'aid' in the traditional sense, but the payment, by a certain group of donors, for the import of environmental services (cleaner air, biodiversity protection, etc.) provided by the South" (Jordan 1995, 306).

Today, the GEF promotes environmentally beneficial projects in developing countries through six focal areas: biological diversity, climate change, international waters, ozone depletion, land degradation, and persistent organic pollutants, as well as multifocal projects (see figure 1). The GEF acts as the financial mechanism for the following global conventions and international agreements: the UN Framework Convention on Biological Diversity, the UN Framework Convention on Climate Change, the Montreal Protocol, the UN Framework Convention on Combating Desertification, and the Stockholm Convention on Persistent Organic Pollutants.

In 2002, the GEF received $2.92 billion for the 2003 6 lending cycle, its largest financial allocation
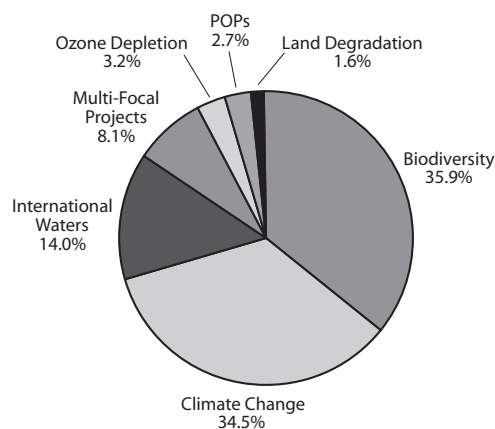
Figure 1

Distribution of GEF allocations by focal area. Source: GEF 2005.

up to that point in time. The United States is the largest single shareholder, supplying 20 percent of GEF financial contributions. However, since the GEF's first funding cycle, the United States has been holding down GEF funding levels. Most recently, it required the GEF to adopt a new resource allocation framework as a condition of continued U.S. financial support (Clémençon 2006).

Tackling Global Environmental Dilemmas and Achieving Results The GEF Evaluation Office independently monitors and evaluates GEF programs and projects. It periodically publishes its assessments in independent overall performance studies (OPS). The 2005 performance study (OPS3) emphasizes the challenge inherent in such evaluations: while some projects have quantifiable benefits and may be evaluated within immediate, short-term timeframes, most GEF projects are large and require long time frames to see results (GEF 2005b). The success of long-term projects hinges not merely on whether the GEF has completed a checklist of environmental goals, but also on whether the project is sustainable; projects must achieve financial sustainability by the time GEF funding ends (Clémençon 2006).

The Ozone Program is notable for its quantifiable, short-term benefits. Although not officially linked to the Montreal Protocol, the GEF has aided countries with economies in transition in achieving their obligations under the protocol. Consumption

of ozone depleting substances (ODS) chlorofluorocarbons, halons, carbon tetrachloride, and methyl chloroform decreased by 99.8 percent from the 1980s to 2003 (GEF 2005b). Much of the success of this small-budget program is due to external factors such as economic slowdowns, but OPS3 finds that the GEF has essentially achieved its goal of eliminating the consumption and emission of ozone depleting substances in countries with economies in transition (GEF 2005b).

Other programs in the GEF portfolio may contribute to environmental benefits, but their success is dependent on long-term project sustainability and will not be quantifiable for decades. Instead, impacts can be measured in terms of socioeconomic and political achievements. The International Waters Program, which has no global convention to follow, has helped to foster open lines of communication and to jump-start regional cooperation among stakeholders and countries, in part by supporting the development of global and regional conventions (Clémençon 2006). Though less quantifiable than projects in the ozone area, the achievements of International Waters programs are nonetheless significant in terms of effecting positive environmental and socioeconomic conditions (GEF 2005b).

Likewise, GEF's energy efficiency portfolio has played an important role "in developing and transforming the markets for energy and mobility in developing countries" (GEF 2005b). The Climate Change Program has seen some socioeconomic and capacity-building impacts, but clear-cut environmental benefits remain blurry: while a particular Climate Change project may bring about carbon reduction, its effects are negligible in view of overall world emissions. One way to think about GEF interventions is that they may only be successful if they are replicable, lead to more funding, or build capacity that makes possible future policy development and implementation (Clémençon 2006). In this light, the success of the Climate Change projects has been limited.

The resources that GEF directs toward preserving biodiversity are notable. However, the impact of these resources remains ambiguous. The GEF's contribution of more than U.S. $2 billion to biodiversity projects from 1991 to 2005 makes it the largest single benefactor of biodiversity (GEF 2005a). The GEF is thus highly influential for the future of the world's biodiversity, but one study reports that it is "difficult, if not impossible, to determine the impact GEF funding has had on the biodiversity it sought to protect" (Horta, Round, and Young 2002, 18). A recent assessment of some 34 biodiversity conservation projects found that in two-thirds of the projects, important outcomes were not sustained after project completion (Dublin, Volonte, and Braun 2004). The GEF's role in achieving the global goal of placing 10 percent of the world's land area in preservation should not be overlooked, and conservation within protected areas has been impressive, but biodiversity loss outside of protected areas continues unchecked (Clémençon 2006).

Since 1994, evaluations of the GEF have emphasized the necessity of a holistic approach that addresses the root causes of environmental problems (Horta, Round, and Young 2002; GEF 2002). The 2005 performance study (OPS3) echoes this call, specifically with attention to sustainability of natural conditions as well as political and social conditions. A biodiversity project, for example, may successfully establish a protected area in the short term, but the sustainability of that area is necessary in order to reduce biodiversity loss in the long term. Likewise, projects in the International Waters area are expected to take decades to establish the institutional and legal conditions to achieve improved water conditions (Gerlak 2004). Attention to the persistence of natural, political, and social conditions is key to the success of these projects.

In addition to sustainability, the 2006 performance study prescribes a more strategic vision and a clearer definition of performance indicators (GEF 2005b). Because of the complex nature of transboundary environmental projects, numerous agencies, organizations, and stakeholders lobby the GEF and a multiplicity of organizations offer guidance. Established in 2003, GEF's strategic priorities were intended to handle the proliferation of guidance coming from various global conventions and con-

ferences but have actually increased guidance and broadened a strategic focus at the GEF (GEF 2005b). OPS3 envisions a synergistic, strategic approach with focused, coherent, and prioritized goals that could capitalize on areas where projects overlap, thus increasing efficiency. An approach that spans the focal areas as well as the various levels of landscape, ecosystem, country, and region will "allow GEF to fulfill its role as catalyst and facilitator of global environmental sustainability" (GEF 2005b, 4).

The halting or reversal of practices leading to the deterioration of the global environment exceeds the resources the GEF can provide. The GEF can, however, provide stimuli to catalyze local and governmental work toward mitigating environmentally harmful effects of various processes. On the learning curve of environmentally beneficial lending, the GEF has climbed quickly, especially within quantifiable, technical areas. While the learning curve at this point flattens out, the topography of complex, systems-oriented programs becomes unstable. The GEF's ability to navigate this terrain is imperative to its survival as the world economy's largest environmental lending organization and to its ability to affect the health of environments worldwide.

See also aid, international; multilateral environmental agreements; trade and the environment

FURTHER READING

Clémençon, Raymond. 2006. "What Future for the Global Environment Facility?" *Journal of Environment and Development* 15 (1): 50 74. Calls for a comprehensive look at the GEF as an institution and a consideration of the funds necessary to achieve its environmental objectives. In light of dwindling funds to the GEF's already modest resource base, this article seeks to stimulate discussion about fundraising and the future of the GEF.

Dublin, H., C. Volonte, and J. Braun. 2004. "GEF Biodiversity Program Study." *Monitoring and Evaluation Unity*. Washington, DC: GEF Secretariat. Program assessment by an independent expert and staff of the GEF Monitoring and Evaluation Unit; a major contribution to OPS3.

Gerlak, Andrea K. 2004. "One Basin at a Time: The Global Environment Facility and Governance of Transbound

ary Waters." *Global Environmental Politics* 4 (4): 108 41. The first independent academic analysis of the International Waters program, this paper reveals broad trends in transboundary water governance, finding that the GEF has fared better with individual projects than with building capacity or enhancing the contractual environment.

Global Environment Facility (GEF). 2002. *Focusing on the Global Environment: The First Decade of the GEF*. Second Overall Performance Study (OPS2). Washington, DC: GEF. Assesses the extent to which the GEF has achieved its main objectives since restructuring in 1994 and policies adopted by GEF Council.

———. 2005a. "Achieving the Millennium Development Goals: A GEF Progress Report." *Global Environment Facility*. Washington, DC: GEF. Juxtaposes GEF environmental objectives with the UN Millennium Development Goals and draws important connections between human and environmental systems.

———. 2005b. "Overall Performance Study: Progressing Toward Environmental Results." Available online at http://www.gefweb.org/MonitoringandEvaluation/MEPublications/MEPOPS/documents/Publications OPS3 E book.pdf. The GEF's third overall performance study launches the fourth replenishment cycle of the GEF and assesses the extent to which the GEF has achieved, or moved toward achieving, its environmental objectives. OPS3 focuses its recommendations on synergistic, integrative strategies that bridge projects across both geographical and institutional levels.

———. 2006. www.gefweb.org. The GEF website provides a rich source for publications and documents by the GEF, as well as descriptions of the GEF's history, organizational scheme, financial contributions, participants, and stakeholders.

Horta, Korinna, Robin Round, and Zoe Young. 2002. "The Global Environment Facility: The First Ten Years Growing Pains or Inherent Flaws?" *Environmental Defense and Halifax Initiative*. Washington, DC: GEF. Evaluates the first ten years of GEF operations, finding it a flawed mechanism in need of fundamental reform. Analyzes the successes and failures of the Small Grants Program.

Jordan, Andrew. 1995. "Designing New International Organizations: A Note on the Structure and Operation of the Global Environmental Facility." *Public

Administration 73: 303 12. Traces the development of the GEF during the pilot phase and restructuring and examines issues likely to dominate a post pilot phase GEF and sketches the broad potential of the GEF.

Sjöberg, Helen. 1994. "From Idea to Reality: The Creation of the Global Environment Facility." *Global Environment Facility*. Washington, DC: GEF. Documents the process leading up to the creation of the pilot phase of the GEF and examines the various proposals made and the negotiations which followed, leading up to the 1994 agreement on the new GEF.

ETHAN C. MYERS AND ANDREA K. GERLAK

■ global imbalances

Global imbalances are large and unsustainable mismatches in crucial macroeconomic variables in major countries or areas of the world economy. Among the most serious and persistent global imbalances in the early 21st century are (1) the large and unsustainable trade deficits of the United States and surpluses of China, Japan, and some other emerging Asian economies, (2) inadequate national savings of the United States and the excessive savings of China and Japan, (3) high unemployment rate of Europe and inadequate growth of Japan, and (4) dismal poverty and slow growth of some of the world's poorest countries in relation to the very high standards of living in advanced countries. Left uncorrected, these global imbalances can lead to serious economic problems in the countries and regions involved and for the world as a whole.

Persistent large U.S. trade deficits could lead to a collapse, or sharp depreciation, of the dollar, which in turn could result in higher interest rates and recession in the United States, and slower economic growth or recession in the rest of the world. Inadequate U.S. savings and excessive savings in China, Japan, and some other (mostly Asian) emerging market economies are leading to large capital inflows from these countries to the United States and a high level of U.S. international indebtedness; but a sudden sharp reduction in such capital inflows could inflict serious economic damage on the United States and

the entire world economy. High unemployment in Europe and slow growth in Japan due to their somewhat rigid or inflexible economies sharply reduce the economic growth of these economies and push them toward trade protectionism in a vain effort to protect labor and unproductive sectors from world competition, especially from China and other dynamic emerging Asian economies. Finally, dismal poverty and slow growth of some of the poorest countries, especially those of sub-Saharan Africa, lead to strife and political unrest in these countries and ethical conflicts and turmoil in the collective conscience of the rich countries.

Trade Imbalances between the United States and East Asia Table 1 shows the enormous and rapidly growing trade deficits of the United States, on the one hand, and the large trade surpluses of China and Japan, on the other hand, since the mid-1990s. The table shows that the U.S. trade deficit rose sharply from $191 billion dollars or 2.4 percent of the U.S. gross domestic product (GDP) in 1996 to $452 billion or 4.6 percent of the U.S. GDP in 2000, and $782 billion or 6.3 percent of U.S. GDP in 2005. Trade deficits in excess of 2 or 3 percent of GDP are deemed by most economists to be unsustainable in the long run and need to be corrected. Table 1 also shows that almost one half of the U.S. trade deficit is with China and the other emerging markets of East Asia and Japan, and it is against the currencies of these nations that the dollar is overvalued. The situation is different with respect to Europe and the euro. Since 2001 the dollar has depreciated substantially with respect to the euro. Overall, the large and rapidly increasing U.S. trade deficits are the result of extremely low and declining savings (table 2) of the United States and its more rapid growth relative to the euro area and Japan over the past decade.

Inadequate U.S. Savings Another serious global imbalance is given by the inadequate savings of the United States, on the one hand, and the excessive savings of China, Japan, and some other emerging Asian economics (sometimes referred to as a global savings glut). Table 2 shows that the savings rate of the United States declined from an average of 16.3

Table 1
Trade imbalances of leading industrial countries and areas, 1996 2005 (in billions of U.S. dollars)

Country/area	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
United States	191.0	198.1	246.7	346.0	452.4	427.2	482.3	547.3	665.4	782.7
US with respect to:										
Western Europe	24.4	24.2	38.3	58.4	74.9	77.1	100.7	116.4	131.5	146.4
Japan	48.7	57.3	65.4	74.8	83.0	70.6	71.8	67.8	77.5	84.7
China	39.6	49.7	57.0	68.7	83.9	83.2	103.2	124.1	162.0	201.7
Other Asia	38.0	41.3	64.8	81.0	104.4	90.2	92.9	100.0	114.8	83.2
Euro area				80.0	28.8	65.2	122.2	120.6	130.4	68.0
Japan	83.6	101.6	122.4	123.3	116.7	70.2	93.8	106.4	132.1	94.0
China	19.5	46.2	46.6	36.0	34.5	34.0	44.2	44.7	59.0	134.2

Source: *Survey of Current Business*, July 2006; IMF *International Financial Statistics Yearbook*, 2006.

percent of GDP in the period 1990 99 to 13.3 percent of GDP in 2005, or by 2.9 percentage points from 1991 to 2005, while U.S. investment rose from an average of 18.7 percent of GDP in the period 1990 99 to 20.0 percent of GDP in 2005, or by 1.5 percentage points from 1991 to 2005. As a result, U.S. net borrowing abroad (negative U.S. net savings) increased from an average of 2.4 percent of GDP in the period 1990 99 to 6.7 percent of GDP in 2005, or by 4.4 percentage points (2.9 1.5 in net savings) from 1991 to 2005. The counterpart of the huge U.S. net borrowing was the large foreign net lending to the United States, primarily by Asian countries, with the euro area practically in savings-investment balance.

Most foreign lending to the United States during the 1990s took the form of foreign direct investments (FDI) attracted by high U.S. growth and profitability. During this period, an increasing amount of foreign capital flows to the United States took the form of foreign official purchases of U.S. government securities and accumulation of dollar reserves, mostly by Japan and Asian emerging market economies, especially China. By buying U.S. government securities and accumulating dollar reserves, Asian countries avoided a large depreciation of the dollar vis-à-vis their own currencies and hence a sharp reduction in their exports to the United States and to the rest of the world (which would result from Asian exports becoming much more expensive for consumers and

businesses paying for them in depreciated dollars). Essentially, Asian countries thus followed an export-led growth strategy financed by lending large sums to the United States. This resulted in a dollar overvalued with respect to most Asian currencies and in a large and unsustainable U.S. trade deficit (the counterpart of the large U.S. net borrowing abroad).

Economists disagree over how long this process can continue. But a significant and rapid reduction in net foreign loans (capital inflows) to the United States could lead to a collapse of the dollar, higher interest rates in the United States, and a U.S. and world economic recession and crisis. There is a broad consensus among economists that the medium- and long-term solution to this problem is for the United States to adopt policies to stimulate domestic savings and curb its expenditures. This is not easy to do because, as most economists would agree, American consumers and firms have become accustomed to living with increasing levels of debt, and it is difficult to change this behavior. An increase in the U.S. savings rate would reduce the need for capital inflows, reduce or eliminate the dollar misalignment, and reduce U.S. trade deficits to sustainable levels. The United States, however, is hardly doing enough to overcome its savings-investment imbalance a problem that could not, in any event, be easily or quickly solved. It has been suggested that China and other emerging market economies should increase domestic expenditures while Japan should restructure

Table 2
Savings, investments, and net savings in major economic areas, 1990 2005 (percent of GDP)

Country/area	Average 1990 99	Average 2000 03	2004	2005	Cumulative change 1991 2005 (percentage points)
United States					
Savings	16.3	15.5	13.4	13.3	2.9
Investment	18.7	19.2	19.6	20.0	1.5
Net savings	2.4	3.7	6.2	6.7	4.4
Euro area					
Savings	21.2	20.9	21.2	20.9	1.0
Investment	19.8	20.7	20.5	20.9	0.7
Net savings	1.4	0.2	0.7	0.0	1.7
Japan					
Savings	31.4	26.7	26.4	26.8	7.6
Investment	29.0	24.0	22.7	23.2	9.7
Net savings	2.4	2.7	3.7	3.6	2.1
China					
Savings	38.7	37.1	46.8	51.3	13.5
Investment	37.0	35.0	43.3	44.1	9.4
Net savings	1.7	2.1	3.5	7.2	4.1

Source: BIS, *Annual Report*, 2006.

its economy to further stimulate its growth. The International Monetary Fund through its new surveillance mandate may encourage the United States, China, and Japan to take cooperative steps to reduce their saving-investment imbalance to sustainable levels.

Structural Imbalances in Europe and Japan
Although savings and investment are in near balance in the euro area (see table 2), this part of the world faces a serious structural imbalance, which has kept its growth rate well below its potential and much lower than the growth rate in the United States since 1996 (see table 3). The restructuring that had taken place in the euro area during the decade from the mid-1990s to the mid-2000s clearly was inadequate, and its economy, especially its labor market, remained excessively rigid by the end of 2007. In addition, some economists argued that the euro area did not pursue the creation of a "new economy" sufficiently aggressively, and it thus was not able to harvest as much benefit as the United States did in the late 20th and early 21st centuries. From 1996 to

2005, the average annual growth rate was 2.0 percent for real (inflation-adjusted) GDP and 0.8 percent for labor productivity in the euro area, as compared with 3.3 percent and 2.2 percent, respectively, in the United States. This has kept the rate of unemployment much higher in the euro area than in the United States and contained European imports over the past decade.

Economists generally agree that the appropriate long-term policy for the euro area to overcome its imbalance would be to accelerate or speed up the restructuring of its economy and liberalize its labor markets, as well as encourage more rapid adoption and spread of new ICT (information and communications technology), and so stimulate labor productivity and growth of its economy. Strong labor opposition, however, has so far prevented European governments from introducing the deep structural reforms necessary. This may be understandable Europe is proud of its high wages and generous social protection benefits, and it seems unwilling to compromise them. It is also difficult to introduce reforms

Table 3
Growth of real GDP and labor productivity, and unemployment rate, 1996 2005 (percentages)

Country/area	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	1996 2005 average
United States											
Real GDP	3.7	4.5	4.2	4.4	3.7	0.8	1.6	2.7	4.2	3.5	3.3
Labor productivity	1.8	2.1	1.9	2.4	1.9	0.9	2.8	2.7	3.1	2.1	2.2
Unemployment rate	5.4	4.9	4.5	4.2	4.0	4.7	5.8	6.0	5.5	5.1	5.0
Euro area											
Real GDP	1.4	2.6	2.7	2.9	4.0	1.9	1.0	0.7	1.8	1.4	2.0
Labor productivity	0.8	1.6	0.7	0.7	1.5	0.3	0.3	0.3	0.9	0.5	0.8
Unemployment rate	10.7	10.6	10.1	9.2	8.2	7.9	8.3	8.7	8.9	8.6	9.1
Japan											
Real GDP	2.6	1.4	1.8	0.2	2.9	0.4	0.1	1.8	2.3	2.7	1.2
Labor productivity	2.1	0.3	1.1	0.6	3.2	0.9	1.4	2.0	2.1	2.3	1.4
Unemployment rate	3.4	3.4	4.1	4.7	4.7	5.0	5.4	5.3	4.7	4.4	4.5

Source: OECD, *OECD Economic Outlook,* June 2006.

and restructure labor markets when the economy is growing slowly. Furthermore, the benefits of re-structuring generally come only over time, but most of the costs are paid up front. The longer the re-structuring is delayed, however, the more the euro area falls behind the United States in real per capita income. In fact, after rapidly reducing the gap in real per capita income from more than 50 percent in 1950 to approximately 10 percent in the 1980s, the euro area average real per capita income slipped back to about 76 percent of U.S. income in 2005, wid-ening the gap to 24 percent.

The Japanese imbalance is also structural and mostly internal in nature, thus requiring, for the most part, domestic policies to correct it. Specifically, Ja-pan was in recession or slow growth for the entire decade of the 1990s, and by the end of 2007 still faced major structural imbalances. Domestic defla-tion rather than international disturbances was the primary cause of the large undervaluation of the yen with respect to the dollar and thus of the large Jap-anese trade surplus vis-à-vis the United States. By most yardsticks, Japan has a unique situation it saves too much and consumes and invests domesti-cally too little (see table 2).

Japan has followed three policies to correct its domestic deflation and structural imbalance. It has pursued a very powerful expansionary monetary policy, which has kept nominal interest rates at a practically zero level (with real rates negative because of domestic deflation). It has adopted an equally powerful expansionary fiscal policy (evidenced by an average annual budget deficit of 6.4 percent of GDP from 1996 to 2005 and 6.8 from 2001 to 2005 and a public debt equal to 172.1 percent of GDP in 2005). Finally, Japan has actively intervened in foreign ex-change markets to prevent a further yen appreciation and a reduction of its exports. In fact, Japan's foreign exchange dollar reserves increased from about $120 billion in 1996 to $830 billion at the end of 2005. Yet, until 2004, Japanese growth remained very subdued. Despite the fact that growth resumed in 2004, remaining structural imbalances still kept Ja-pan performing below its potential by 2007.

Dismal Poverty and Growing Global Inequality
Still another serious global imbalance in the world today is dismal poverty and slow growth of some of the world's poorest countries, especially those of sub-Saharan Africa. Table 4 gives the population and the per capita income of various countries or groups of

Table 4

Population and economic and health indicators, 1990 2004

Country/region	Population in 2004 (millions)	Income per capita		Infant mortality rate per 1,000 live births		Life expectancy at birth (years)	
		Dollars 2004	Growth rate 1990 2004 (% per year)	1990	2004	1990	2004
Low and middle income	5,344	1,502	2.2	69	59	63	65
Sub Saharan Africa	719	600	0.3	111	100	49	46
East Asia and Pacific	1,870	1,416	6.4	43	29	67	70
of which, China	1,296	1,500	7.0	38	26	69	71
South Asia	1,448	594	3.6	86	66	59	63
of which, India	1,080	620	4.2	80	62	59	63
Europe and Central Asia	472	3,295	0.5	40	29	69	69
Middle East and N. Africa	294	1,972	1.1	60	44	64	69
Latin America and Caribbean	541	3,576	1.1	43	27	68	72
High income economies	1,001	32,112	2.6	9	6	76	79
World	6,345	6,329	0.5	64	54	65	67

Source: World Bank, *World Development Report*, 2006; and *World Development Indicators*, 2006.

countries in 2004, as well as the growth in real per capita income from 1990 to 2004, and infant mortality and life expectancy in 1990 and 2004. The table shows that the average per capita income of all developing economies and former communist countries was only $1,502 in 2004 ($620 and $1,500 for India and China, respectively) as compared with $32,112 in high-income developed economies. Worse still, the average growth of real per capita income was close to zero in sub-Saharan Africa (as a result of drought, wars, rapid population growth, the spread of HIV, and the general failure of development efforts), only 0.5 percent in Europe and Central Asia (because of economic restructuring after the collapse of communism), and 1.1 percent in the Middle East (because of wars, political turmoil, and the sharp decline in petroleum prices during the 1990s).

The average growth of real per capita income was also very low (only 1.1 percent) in Latin America and the Caribbean between 1990 and 2004 because of political turmoil and failure in the development effort. Only in East Asia and the Pacific economies (and in particular, in China) did the real per capita

income increase very rapidly from 1990 to 2004. In South Asia, the growth of real per capita income, while not as spectacular as in East Asia, was very respectable, largely fueled by the economic revitalization of India. The table also shows that infant mortality is much higher and life expectancy much lower in low-income developing countries, especially in sub-Saharan Africa, than in high-income developed countries, but major improvements were made in both measures throughout the world from 1990 to 2004 (except, again, in sub-Saharan Africa). But the wide disparities in per capita incomes between rich and poor countries, especially the poorest developing countries, can be regarded as one of the most serious global imbalances in the world economy today.

An international economic system that has spread the benefits from globalization so unevenly can hardly be said to be functioning properly   not to mention equitably. And a world where millions of people starve not only is unacceptable from an ethical point of view, but also can hardly be expected to be peaceful and tranquil. The huge difference in stan-

dards of living between rich countries and the poorest ones certainly represents one of the most serious global imbalances in the world economy today.

See also balance of payments; Bank of Japan; conflicted virtue; dollar standard; European Central Bank; foreign exchange intervention; international financial architecture; International Monetary Fund surveillance; international reserves; liquidity trap, the; mercantilism; poverty, global; twin deficits

FURTHER READING

Bank for International Settlements (BIS). 2006. *Annual Report.* Basel: BIS. Report by the Bank for International Settlements on the world's financial and banking situation.

Campano, Fred, and Dominick Salvatore. 2006. *Income Distribution.* New York: Oxford University Press. Discusses how to measure changes in income distribution and income inequalities within and among nations.

de Rato, Rodrigo. 2005. "Global Imbalances and Global Poverty Challenges for the IMF." Available at http://www.imf.org/external/np/speeches/2005/022305.htm. An analysis of the relationship between global imbalances and global poverty by the former managing director of the International Monetary Fund.

Eichengreen, Barry. 2004. "Global Imbalances and Lessons of Bretton Woods." NBER Working Paper No. 10497 (May). Washington, DC: National Bureau of Economic Research. Shows how global imbalances led to the collapse of the Bretton Woods System in the early 1970s and what can we learn from it.

Geitner, Timothy. 2006. "Policy Implications of Global Imbalances." Remarks at the General Financial Imbalances Conference at Chatham House, London (January 23). Available at http://www.ny.frb.org/newsevents/speeches/2006/gei060123.html. Proposals for correcting the global imbalances facing the world toward the end of the first decade of the 21st century by the president of the Federal Reserve Bank of New York.

International Monetary fund (IMF). 2005. "Global Imbalances: A Saving and Investment Perspective." *World Economic Outlook* (September). Washington, DC.: IMF, 91 124. Provides data on the saving investment imbalance in the world economy and proposals for how to overcome it.

. 2006a. *Annual Report.* Washington, DC: IMF. An analysis of the world's economic and financial situation by the International Monetary Fund.

. 2006b. *World Economic Outlook* (September). Washington, DC: IMF. Provides a biyearly overview of the world economy by the International Monetary Fund.

OECD. 2006. *Economic Outlook* (June). Paris: OECD. A biyearly overview of the world economy by the Organisation for Economic Co operation and Development.

Rajan, Raghuram G. 2005. "Global Imbalances An Assessment." http://www.imf.org/external/np/speeches/2005/102505.htm. An evaluation of global imbalances facing the world toward the end of the first decade of the 21st century by the former chief economist of the International Monetary Fund.

Rogoff, Kenneth. 2006. "Global Imbalances and Exchange Rate Adjustment." *Journal of Policy Modeling* (August 2006): 695 99. An examination of how global imbalances could be eliminated or reduced by a coordinated approach by the world's major economies by the former chief economist of the International Monetary Fund.

Salvatore, Dominick. 2005. "Currency Misalignments and Trade Asymmetries among Major Economic Areas." *The Journal of Economic Asymmetries* 2 (1): 1 24. Shows how misalignments among the currencies of the major economies are the results of global structural imbalances and how to correct them.

. 2006. "Twin Deficits in the G 7 Countries and Global Structural Imbalances." *Journal of Policy Modeling* (August): 701 12. Shows the relationship between budget deficits and current account deficits in the leading industrial economies.

United Nations. 2006. *Human Development Report.* New York: UN. Provides different measures of the standard of living of member nations by the United Nations.

World Bank. 2006a. *Annual Report.* Washington, D.C.: World Bank. An analysis of the world's economic and financial situation by the World Bank.

. 2006b. *World Development Indicators.* Washington, D.C.: World Bank. A wealth of data on each country's economic and financial situation by the World Bank.

DOMINICK SALVATORE

■ global income inequality

Prevailing concerns about economic inequity in the world reflect many aspects of living standards and how they are distributed, and no single measure can hope to capture all those concerns. As conventionally measured, "inequality" and "poverty" are quite different aspects of income distribution, in that the former focuses on the (absolute or relative) disparities in income (or consumption), while the latter focuses on absolute levels of deprivation, which depend on the average levels of living in society as well as inequality.

Both poverty and inequality are closely related to a third concept, "social welfare," which aggregates welfare levels, which are taken to depend, at least in part, on income. (Differences in "income needs" may also play a role.) Poverty and social welfare are both sensitive to income changes at the bottom of the distribution: if everything else is held constant, as poverty rises, social welfare falls. They differ, however, in that standard measures of social welfare are also sensitive to changes at the top of the income ladder, whereas poverty measures typically are not. Inequality also relates to social welfare, since most social welfare functions have curvature properties that penalize higher inequality at any given mean income (Atkinson 1970).

Well-being itself has many dimensions, of course, and interpersonal aggregation is fraught with difficulties. Some people argue that the standard of living should be assessed in terms of capabilities   the set of "beings and doings" one is capable of enjoying rather than income or consumption (see, for example, Sen 1985). This can be interpreted as an issue of how best to deal with interpersonal heterogeneity in income "needs."

There is not enough room in this entry to do justice to all these issues. The focus here is on global consumption and income distribution, and how they are changing.

Concepts, Methods, and Data Conceptual and methodological differences lie at the root of the ongoing debates about global inequality. There are significant differences in four main areas: the welfare indicator ("inequality of what?"), the definition of the distribution, the choice of inequality measure, and the appropriate data sources. Without some understanding of these differences, it is impossible to take an informed position in this debate.

Inequality of What? Standard practice has been to rank households by consumption or income per person. The consumption or income numbers that can be formed from most nationally representative large household surveys are reasonably comprehensive, in that they span the commodity space or all income sources (with imputed value for income in kind from production for one's own use or consumption). But they can hardly be considered complete metrics of welfare; access to subsidized health care or schooling is generally excluded, for example. Hence it is important to also look separately at key "nonincome" dimensions of welfare that capture the missing components. We will return to this topic, but for now we focus on the standard consumption or income measures available from surveys.

The choice between consumption and income is not a minor one, since inequality is generally lower for consumption than for income, due to consumption smoothing (whereby savings or borrowing are used to assure that living standards do not vary as much over time as incomes do). The consensus in the development literature is that consumption is both the conceptually preferable measure   given some degree of smoothing in the presence of intertemporal income variability   and more accurately measured in surveys. (Against this view, however, some scholars have argued that income may better reflect opportunities for consumption.)

The near-universal normalization by household size is also questionable. For example, it does not allow for economies of size in consumption (whereby two people can live more cheaply together than apart), and doing so can affect inequality and poverty comparisons, such as the claims often heard that larger households are poorer (Lanjouw and Ravallion 1995).

Naturally, household surveys measure expenditures in local currencies, which then have to be converted to a single currency. The basic choice is between market exchange rates, which measure in-

ternational purchasing power in terms of traded goods, and purchasing power parity (PPP) exchange rates, which adjust for the fact that nontraded goods tend to be cheaper in poorer countries. Almost all researchers prefer PPPs, although there is more than one way to measure PPP exchange rates. The Geary-Khamis (GK) method used by the Penn World Tables (PWT) underestimates global inequality since the quantity weights used to compute the international price indexes give greater weight to consumption patterns in richer countries, which in turn results in an overestimate of expenditures in poorer countries. The Eltöte, Köves, and Szulc (EKS) method, which is the multilateral extension of the bilateral Fisher index, attempts to correct for this bias. (On the differences between the GK and EKS methods and implications for global poverty measures, see Ackland, Dowrick, and Freyens 2006. Since 2000, the World Bank's global poverty and inequality measures have been based on the Bank's PPPs, which use the EKS method.)

Although these are all difficult measurement issues, the main sources of the differences in existing numbers on the evolution of global poverty and inequality lie elsewhere.

What Distribution? There are three main — and very different — ways of defining the "world income distribution." The first is the distribution of country mean incomes, unweighted by population; we call this the "intercountry income distribution." This is of interest primarily to those who view the country, rather than the individual, as the main unit of interest. If gross domestic product (GDP) per capita is used as a measure of the country mean income, this concept corresponds closely to the macroeconomic literature on testing for international income convergence (see, e.g., Quah 1996; Pritchett 1997). The second distribution is that of country means, weighted by population — the "international income distribution." This places higher weight on more populous countries.

The third definition focuses on the distribution of *individual* incomes, across all people in the world. This is the "global income distribution," and for those concerned with the individual — rather than

the country — as the fundamental unit of analysis, this third concept is clearly preferable. If inequality between rich Chinese (or Americans) and poor Chinese (or Americans) is rising, why should this not be counted as part of world inequality? The earlier literature focused on intercountry and international inequalities largely for data availability reasons. It is only since about the mid-1980s that household survey data have become available for a sufficient number of countries so as to enable a computation of global inequality. (These three measures — intercountry, international, and global — correspond, respectively, to inequality concepts 1, 2, and 3, as defined by Milanovic 2005. For decomposable measures, international income inequality corresponds to the between-country component of global inequality.)

What measure? A key conceptual distinction too often ignored in applied work — is between *relative* and *absolute* measures of inequality. We consider relative measures first. One of the most popular inequality measures is the Gini index, which is constructed by summing income differences, expressed as a ratio to the mean, across all individuals in a population and then dividing it by the number of people. This measure has an intuitive interpretation: if two individuals are drawn at random, then the expected difference in their incomes, normalized by the mean, will be equal to twice the Gini index. The index can also be depicted graphically. If one orders the population from poorest to richest, and graphs the cumulative income share against the cumulative population share of each percentile (thereby obtaining the "Lorenz curve"), then the Gini index is twice the area between the Lorenz curve and the diagonal. There are many other inequality measures; a good overview can be found in Cowell (2000). Another popular measure is the mean log deviation, given by the average proportionate gap between mean income and actual income; this measure has some attractive properties, including decomposability into (population-weighted) between-group and within-group components (Bourguignon 1979).

These measures of inequality are designed to be *scale-invariant*, meaning that they do not change if

every income in the distribution is multiplied by a positive constant. The resulting measure then depends solely on income *ratios*, in contrast to absolute measures, which depend on the absolute *differences* in incomes. If every person becomes 10 percent richer, then a relative measure of inequality does not change, even though the (absolute) gains to the rich are many times larger than those to the poor.

Many commentators on globalization appear to focus on absolute income differences rather than relativities. Being sensitive to differences rather than ratios, an absolute inequality measure would increase in our hypothetical scenario of a 10 percent rise in everyone's incomes. Though not often used in practice, an absolute version of the Gini index can be constructed by dropping the normalization of income differences by the mean, so that the measure is one-half of the mean difference in incomes of pairs of individuals.

It may well be the case that much of the debate about what is happening to inequality in the world is actually a debate about whether one thinks about "inequality" in relative or absolute terms (Ravallion 2004). Yet the choice between absolute and relative inequality measures is not a matter of right or wrong. They correspond to different procedures for aggregating gaps between incomes, and thus to different normative concepts of inequality. (Ravallion 2004 discusses this issue further and the implications of other conceptual distinctions, such as between vertical and horizontal inequality, for one's assessment of the impact of globalization on inequality.)

Other observers in the globalization debate focus more on absolute levels of living of the poor, rather than the gaps (whether absolute or relative) between the rich and poor. The measures of poverty found in practice reflect the absolute consumption or income shortfalls from some agreed minimum standard, the poverty line, which aims to have the same real value over time and (when relevant) space. Suppose again that all incomes increase by 10 percent. Relative inequality will remain the same and absolute inequality will have risen. But the incidence of absolute poverty will have fallen, reflecting the higher living standards of the poor. Although this entry focuses

mainly on global inequality, this would give an incomplete picture if we did not also look at what has been happening to global poverty.

What data sources? One might have guessed that there was general agreement on data sources, but that is not the case, and the choices made do matter. There are essentially two approaches. The first relies on secondary data sources, from existing data compilations such as the World Bank's *World Development Indicators*, and the second relies more heavily on primary sources, including unit-record data from household surveys.

A number of studies have used quintile shares or inequality measures derived from various inequality data bases as estimates of inequality in each individual country (examples include Bhalla 2002; Sala-i-Martin 2006). These shares are often derived from already grouped data (rather than from microdata), and they refer to different welfare concepts (income or consumption) and different recipient units (households or individuals). In order to increase the data coverage, the quintiles are pooled together without due attention to their comparability.

Another data issue is the choice between using national accounts (NAS) versus household survey means as the appropriate mean for the distribution. While GDP per capita numbers computed from NAS were traditionally used to construct intercountry and international distributions, household surveys have been the traditional tool for measuring poverty and inequality. The aggregates from these sources generally do *not* agree, and that is hardly surprising when one probes the way the numbers are generated. It seems that over relatively long time periods the two data sources tend to converge to a common ratio in most countries, but the remaining difference in levels can be sizable, and there are some notable short-term divergences (as documented by Ravallion 2003 and Deaton 2005).

Whether using quintile estimates from secondary data or original microdata, some authors rescale individual incomes in each country so as to have a mean corresponding to GDP per capita (from the NAS). They often argue that surveys do not capture a number of determinants of welfare, such as publicly

provided goods. They also worry about income underreporting in these surveys.

There are problems with this rescaling method. On the one hand, it is not clear that the NAS data can provide a more accurate measure of mean *household* welfare than the survey data that were collected precisely for that purpose. On the other hand, even acknowledging the problems of income underreporting and selective survey compliance, there can be no presumption that the discrepancies with the NAS are distribution neutral; more plausibly, the main reasons why surveys underestimate consumption or income would also lead to an underestimation of inequality. For example, Banerjee and Piketty (2005) attribute up to 40 percent of the difference between the (higher) growth of GDP per capita and (lower) growth of mean household per capita consumption from household surveys in India to unreported increases in the incomes of the rich. Selective compliance with random samples could well be an equally important source of bias, although the sign is theoretically ambiguous; Korinek, Mistiaen, and Ravallion (2006) provide evidence on the impact of selective nonresponse for the United States.

Nor are the NAS always reliable. For instance, global inequality results depend crucially on China, but China's GDP per capita numbers are the subject of intense debate. Maddison (2003) and, to a lesser extent, Penn World Tables give significantly higher values for China's GDP per capita in the 1950s and 1960s than the official Chinese statistics. The differences have important implications for the estimation of global inequality.

For these reasons, there is a growing recognition that adjustment of household survey data by the NAS mean is probably undesirable (Deaton 2005; Bourguignon and Morrisson 2002; Anand and Segal 2008). The best current practices in measuring global inequality and poverty increasingly rely, to the maximum extent feasible, on nationally representative household survey data.

The Evolution of Global Inequality and Poverty Acknowledging these conceptual, data, and methodological caveats, what can be said about trends in the world distribution of income?

Bourguignon and Morrisson (2002) constructed a time series of world inequality estimates for the period from 1820 to 1992. For all but the last ten to twenty years of that series, disaggregated household survey data are not available for a large number of countries. Countries were grouped into 33 blocks, with block composition changing over time, depending on data availability (for details, see Bourguignon and Morrisson 2002). The distributions are constructed in such a manner that all the members of a block are assumed to have the same distribution as a country for which distribution data are available. The authors construct a distribution based on decile (and some ventile) shares, and on GDP per capita figures. Individuals are assumed to have the same incomes within tenths (or twentieths) of the distribution, where that income corresponds to the group's share of GDP per capita. This hybrid of the concepts discussed in the previous section was the compromise that allowed the authors to construct a long time series covering most of the 19th and 20th centuries. Given the long-run perspective of this exercise, however, it is likely that some of the problems discussed in the previous section were only moderately important. In particular, the estimated evolution of GDP per capita over such a long period is likely to be very strongly correlated with any measure of household welfare.

The main finding of the study is that world inequality rose almost continuously from the onset of the industrial revolution until the First World War. During that period, the Gini coefficient rose from 0.50 to 0.61. Although inequality was also rising within most countries for which data were available, the real driving force for this increase in global disparity was inequality *between* countries, that is, international inequality (see figure 1).

Between the two World Wars, and until around 1950, a decline in within-country inequality was observed, but the rise in inequality across countries continued apace and proved to be the dominant force. The world Gini coefficient rose further to 0.64. From the middle of the 20th century on, the rise of global inequality slowed, as Japan and parts of East Asia started growing faster than Europe and North
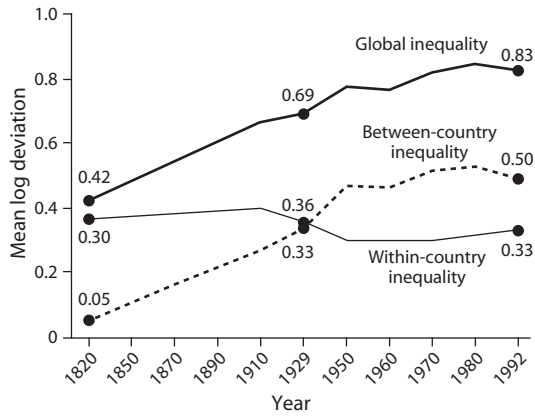
Figure 1

Inequality between and within countries. Source: Authors' manipulation of data from Bourguignon and Morrisson (2002).

America. This process became particularly pronounced after the takeoff of China in the 1980s. Broadly speaking, global inequality changes in the second half of the 20th century were much less significant than in the previous 130 years: there was certainly a reduction in the growth of inequality and, toward the end of the period, inequality may have started to decline.

When considering the last decades of the 20th century, however, better and more comprehensive data are available. Household survey data coverage, and the data's availability to researchers, increased dramatically in the 1980s and 1990s, and it became possible to construct not only intercountry and international inequality series based on a broader set of countries, but also real global income distributions, from the microdata.

Looking at the second half of the 20th century with these data, Milanovic (2005) and World Bank (2006) highlight two interesting regularities. First, even as (unweighted) intercountry inequality continued to grow between 1950 and 2000, international inequality began to fall. The disparate behavior in these two inequality concepts has been one of the reasons behind disagreements on globalization and inequality. The continuing rise in intercountry inequality (to which Pritchett 1997 refers as "divergence, big time") was due largely to slow growth in most poor (and small) countries, relative to some middle-income and richer countries. The decline in international inequality, which refers to a population-weighted distribution, was due fundamentally to rapid growth in two giant countries that started out very poor: China and, to a lesser extent, India. As figure 2 suggests, once China and India are excluded from the international distribution, the post-1980 trend in that inequality concept changes dramatically and becomes much closer to the rising trend in intercountry inequality.

The second regularity is that the last two decades of the 20th century saw a resumption in the upward trajectory of within-country inequality, defined as the aggregate contribution of within-country inequality to total inequality. The rise in within-country inequality prevented the decline in international inequality (which began, slowly, around the 1960s) from translating immediately into a decline in global inequality. Recall that global inequality is the sum of (appropriately aggregated) within-country inequality and international inequality. Indeed, Milanovic (2002, 2005) finds that global income inequality between people was still rising between 1988 and 1993, but appears to have fallen between 1993 and 1998. This is confirmed by World Bank (2006), which extends Milanovic's data set by a couple of years.
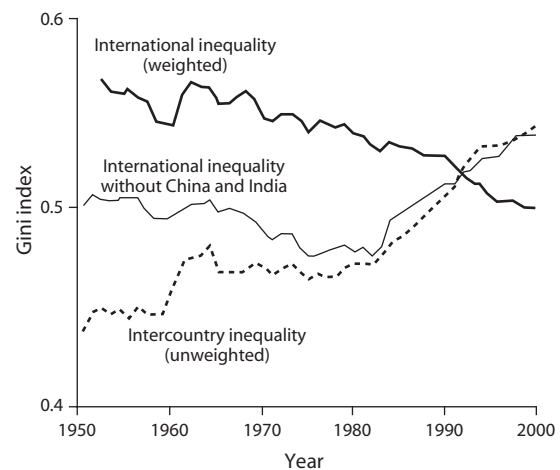
Figure 2

Intercountry inequality and international inequality. Source: Milanovic (2005).

The foregoing discussion has been about relative inequality. What about the competing concept of absolute inequality, which depends on the absolute gaps in levels of living between the "rich" and the "poor"? As figure 3 shows, the two concepts give rise to completely different trends for international inequality: whereas all relative inequality measures shown fall from 1990 on, all absolute measures record substantial increases. The difference is practically as important when considering global inequality. It would slightly increase and plateau when defined in relative terms and increase drastically when defined in absolute terms.

Recall that none of these measures tell us directly about the absolute standard of living of poor people. Has rising inequality reflected falling living standards for the world's poor? Using the longest available period of time with consistent series — 1981–2004 — the World Bank's latest estimates (reported in Chen and Ravallion 2007) show that the poverty rate for the developing world as a whole fell from 40 percent in 1981 to 18 percent in 2001, judged by the $1/day standard at PPP (figure 4). Here too measurement methods matter. Using secondary sources on inequality and NAS data on mean income, Sala-i-Martin (2006) finds that the global poverty rate fell from 13 percent in 1980 to 7 percent in 1998. By both methods, the poverty rate was almost halved over 1980–2000, although the levels are quite dif-

Figure 4

Incidence of absolute poverty in the developing world. This figure gives the estimated percentage of the population of the developing world living below $32.74 per month ("$1 a day") and $65.48 per month ("$2 a day") at 1993 consumption PPP. The lines u + r give the corresponding percentages when one incorporates an allowance for the higher cost of living in urban areas. Source: Chen and Ravallion (2007).

ferent. However, GDP includes much more than household consumption, and it is not clear why one would use the same poverty line when switching from household consumption from surveys to GDP from national accounts. Sala-i-Martin's estimates based on a $2/day line — to allow for the non household consumption share of GDP — accord quite closely with the World Bank's estimates using $1/day.

While there is broad agreement that the world is making progress against absolute poverty, there are some important regional exceptions, notably much of sub-Saharan Africa over the 1980s and 1990s. It is also notable that a sizable share of the overall progress has been due to the success against poverty of just one country, China (Chen and Ravallion 2007).

Inequality in Other Dimensions Although this entry (and the broader debate) has focused on income inequality trends, there should be no presumption that it is the only inequality that matters. Indeed, from some perspectives, international disparities in health status and educational achievement may matter just as much (in addition to being instrumentally important in shaping income inequality and poverty). Since around 1930 there has

Figure 3

Absolute and relative inequality in the world. Source: Atkinson and Brandolini (2004).

been convergence in the intercountry and international distributions of life expectancy at birth (LEB). As mean world (weighted) LEB rose from 53.4 years in 1960 to 64.8 years in 2000, its distribution moved from bimodality to unimodality and the coefficient of variation fell from 0.233 to 0.194 (World Bank 2006). This heartening trend was partly reversed, however, during the 1990s, when LEB fell precipitously in some of the world's poorest countries, due largely to the spread of HIV/AIDS.

Educational inequality, measured by the distribution of years of schooling, has also fallen substantially since the 1960s. As mean years of schooling in the world rose from 3.4 in 1960 to 6.3 in 2000, the coefficient of variation fell from 0.739 to 0.461. (Note that inequality measures for variables like life expectancy or years of education have to be interpreted with care. Both variables are close to being bounded from above, and inequality tends to fall automatically when the mean increases.) This pattern of rising means and falling inequality in attainment was common to all regions of the world and, in addition, all regions also saw a reduction in gender disparities, as measured by the male to female schooling ratio (World Bank 2006).

Unfortunately, this reduction in *attainment* inequality has not always meant a reduction in the disparities in true educational *achievement*. Indeed, internationally comparable test score data suggest that these disparities remain strikingly large with, for example, the reading competence of the average Indonesian student in 2001 being equivalent to that of a student in the seventh percentile of the French distribution.

These changes in the distribution of health and education should be taken into account when assessing global inequality in a broad sense. While this entry provides only a very brief summary of the existing evidence along each dimension, a number of scholars have attempted to explore the correlations among the different dimensions. Because increases in longevity have been greater in poorer countries, for instance, Becker, Philipson, and Soares (2005) argue that inequality in measures of well-being that account for quantity, as well as quality, of life have been declining throughout the postwar period.

Another important aspect is the correlation of incomes over time, and the lack of mobility of countries in the international distribution. The persistence of poverty and income gaps across countries or individuals is not adequately represented by measures that are based on static snapshots of the global or international distribution of income. Relative income dynamics matter. Somehow, the severity of inequality, as observed at two different points of time, depends on whether individuals keep the same position or whether they switch. This is the reason for the attention given to the issue of convergence and the accompanying concept of intercountry inequality; for further discussion see Bourguignon, Levin, and Rosenblatt (2004), who analyze in some detail this issue of country mobility in the international distribution.

More work is needed to properly evaluate the extent of "global inequity" in all its dimensions and its evolution over time. Yet, despite the conceptual and methodological minefield inherent in measuring world income inequality and poverty, it is possible to reach agreement on some key stylized facts:

- Global income inequality was high in the early 19th century, but is even higher today.
- From the industrial revolution until roughly the middle of the 20th century, inequality rose in all three concepts of world income distribution: intercountry, international, and global.
- From around 1950 onward, although intercountry inequality continued to rise, (population-weighted) international inequality first stabilized and then declined. This reflects the catch-up of Asia with Europe and North America.
- Rising within-country inequality has attenuated the decline in global (interpersonal) inequality associated with growth in some of the poorest and largest countries, notably China and India. Nevertheless, global inequality appears to have fallen, or at least reached a high plateau, during the late 1990s.
- Although world inequality in the first few years of the 21st century has not yet been

properly analyzed, it seems likely that this trend has continued. Since 2002, the mean growth rate for low-income countries has been above the average rate for high-income countries, for the first time since the 1960s.

- Absolute income inequality in the world—the absolute gap between "rich" and "poor"—has been rising since at least 1970 and probably for a long time prior to that.
- Even so, the incidence of absolute poverty in the world as a whole has been falling since at least the early 1980s, though more rapidly in some periods and regions than others.
- Some large disparities in human development persist, although it is encouraging that global inequalities in health and education attainments have been falling overall. The scourge of HIV/AIDS has threatened this progress, with falling life expectancies in some of the poorest countries.

See also international income convergence; poverty, global

FURTHER READING

Ackland, Robert, Steve Dowrick, and Benoit Freyens. 2006. "Measuring Global Poverty: Why PPP Methods Matter." Mimeo. Research School of Social Sciences, Australian National University. Considers how differences in PPPs used affect calculation of global inequality.

Anand, Sudhir, and Paul Segal. 2008. "What Do We Know about Global Income Inequality?" *Journal of Economic Literature* 46 (1): 57 94. A critical assessment of methods and results of global inequality calculations.

Atkinson, Anthony B. 1970. "On the Measurement of Inequality." *Journal of Economic Theory* 2 (3): 244 63. Explores measures of inequality and their welfare implications.

Atkinson, Anthony, and Andrea Brandolini. 2004. "Global Income Inequality: Absolute, Relative, or Intermediate?" Paper presented at the 28th General Conference of the International Association for Research on Income and Wealth. August 22. Cork, Ireland. How standard results are overturned if instead of measures of relative inequality one uses absolute inequality.

Banerjee, Abhijit, and Thomas Piketty. 2005. "Top Indian Incomes, 1922 2000." *World Bank Economic Review* 19 (1): 1 20. Long term evolution of income share of the very rich in India based on tax returns.

Becker, Gary, Tomas Philipson, and Rodrigo Soares. 2005. "The Quantity and Quality of Life and the Evolution of World Inequality." *American Economic Review* 95 (1): 277 91. How improvements in life expectancy and their lower dispersal across countries might imply less inequality across global citizens measured over their lifetime.

Bhalla, Surjit. 2002. *Imagine There's No Country: Poverty, Inequality, and Growth in the Era of Globalization.* Washington, DC: Institute for International Economics. Both global poverty and inequality are argued to be much lower than conventionally thought.

Bourguignon, François. 1979. "Decomposable Income Inequality Measures." *Econometrica* 47 (4): 901 20. Why decomposability (between recipients) matters for inequality measures and what are the measures that satisfy it.

Bourguignon, François, Victoria Levin, and David Rosenblatt. 2004. "Declining International Inequality and Economic Divergence: Reviewing the Evidence through Different Lenses." *Economie Internationale* 100: 13 25. Reviews the reasons why rigorous statistical analysis based on different definitions leads to contradictory conclusions about the evolution of global inequality.

Bourguignon, François, and Christian Morrisson. 2002. "Inequality among World Citizens: 1820 1992." *American Economic Review* 92 (4): 727 44. Calculates long term trends in global inequality, both of income and income adjusted for life expectancy among countries.

Chen, Shaohua, and Martin Ravallion. 2007. "Absolute Poverty Measures for the Developing World, 1981 2004." Policy Research Working Paper No. 4211. Washington, DC: World Bank. Estimates poverty head counts using microdata from household surveys.

Cowell, Frank. 2000. "Measurement of Inequality." In *Handbook of Income Distribution*, vol. 1, edited by Anthony B. Atkinson and François Bourguignon. Amsterdam: North Holland, 87 166. Measures of inequality, their characteristics, and the connection between inequality and welfare measurement.

Deaton, Angus. 2005. "Measuring Poverty in a Growing World (or Measuring Growth in a Poor World)." *Review*

of Economics and Statistics 87 (2): 353 78. Issues in methodology of poverty and inequality measurement, especially the choice between survey derived and na tional accounts mean income or consumption.

Korinek, Anton, Johan Mistiaen, and Martin Ravallion. 2006. "Survey Nonresponse and the Distribution of Income." *Journal of Economic Inequality* 4 (2): 33 55. How lower response among rich households biases measures of inequality and poverty.

Lanjouw, Peter, and Martin Ravallion. 1995. "Poverty and Household Size." *Economic Journal* 105 (433): 1415 35. How to correctly account for household economies of size and scale when measuring poverty.

Maddison, Angus. 2003. *The World Economy: Historical Statistics.* Paris: OECD Development Centre Studies. GDP and population statistics, often estimated and interpolated, for practically all countries of the world, and descriptions of main long term trends of world economy.

Milanovic, Branko. 2002. "True World Income Distribu tion, 1988 and 1993: First Calculation Based on Household Surveys Alone." *Economic Journal* 112 (476): 51 92. Calculation of global inequality based mostly on microdata from household surveys.

———. 2005. *Worlds Apart: Measuring International and Global Inequality.* Princeton, NJ: Princeton University Press. How intercountry, international, and global in equality have evolved after the end of World War II up to the year 2000.

Pritchett, Lant. 1997. "Divergence, Big Time." *Journal of Economic Perspectives* 11 (3): 3 17. Intercountry income inequality was much less at the time of the industrial revolution than in the late 20th century.

Quah, Danny. 1996. "Empirics for Economic Growth and Convergence: Stratification and Convergence Clubs." *European Economic Review* 40 (6): 427 43. Discusses the shape of intercountry distribution of mean incomes under different scenarios regarding income convergence or divergence.

Ravallion, Martin. 2003. "Measuring Aggregate Economic Welfare in Developing Countries: How Well Do Na tional Accounts and Surveys Agree?" *Review of Economics and Statistics* 85 (3): 645 52. Explores differences in the time path of the means calculated from household sur veys and national accounts.

———. 2004. "Competing Concepts of Inequality in the Globalization Debate." In *Brookings Trade Forum 2004*, edited by Susan Collins and Carol Graham. Wash ington, DC: Brookings Institution, 1 38. How focus on absolute or relative measures of inequality affects our view of globalization.

Sala i Martin, Xavier. 2006. "The World Distribution of Income: Falling Poverty and . . . Convergence. Period." *Quarterly Journal of Economics* 121 (2): 351 97. Com bines GDPs per capita and distributional data to make estimates of global poverty and inequality.

Sen, Amartya. 1985. *Commodities and Capabilities.* Am sterdam: North Holland. Explores how welfare can be defined in terms of capacity to perform certain functions rather than solely in terms of access to goods.

———. 2001. "Ten Theses on Globalization." *New Per spectives Quarterly* 14 (4): 9 15. Inequality is at the core of the globalization debate even if globalization is not a zero sum game.

Sutcliffe, Bob. 2003. "A More or Less Unequal World? World Income Distribution in the 20th Century." Working Paper Series No. 54. University of Massa chusetts Amherst, Political Economy Research Institute. Published as "World Inequality and Globalization." *Oxford Review of Economic Policy* 20 (1): 15 37. A crit ical review of several studies of global inequality.

World Bank. 2006. *World Development Report: Equity and Development.* New York: Oxford University Press. As sesses global inequality in income, education, and health outcomes, and places it into the context of inequality of opportunity.

FRANÇOIS BOURGUIGNON, FRANCISCO FERREIRA, BRANKO MILANOVIC, AND MARTIN RAVALLION

■ **global production networks (GPNs)**

See commodity chains

■ **global public goods**

Whether people live in industrial or developing countries, whether they are poor or rich, their well-being, in large measure, depends on the availability of

a well-tuned balance between private goods, like food or clothing, and public goods, like peace and security, rule of law, communicable disease control, or financial stability. Many public goods have assumed a global dimension. They are global in that their benefits (and, as the case may be, their costs) span national borders, making an impact on countries in several regions and affecting not only present but perhaps also future generations.

Definition of Public Goods and Global Public Goods The classification of goods as private or public is based on their consumption properties. Goods that are made or left excludable and for which clear property rights can thus be assigned are referred to as private goods. It is up to its owner to decide how and by whom a private good can be used or consumed.

In contrast, public goods are nonexclusive in consumption. They are in the public domain, available for all. The term *good* in this context has no value connotation. People's preferences for these goods may vary widely, due to socioeconomic, political, cultural, and geographic factors. Some may perceive a public good such as financial stability as a top priority; and others may accord greatest importance to exploring new, drought-resistant crop varieties. Yet all face the same public goods, at the same provision level. As a result, the question of which public goods to provide, at which quantity, in what form, and at what net cost or net benefit to whom is often highly contested.

Moreover, in many perhaps even most instances, "publicness" and "privateness" do not constitute an innate property of the good. Rather, a good may be in the public domain, available for all to consume, for various reasons, including the following:

- Infeasibility of exclusion for technical or economic reasons (as in the case of the moonlight);
- Deliberate policy choice (as in the case of norms and institutions such as the regime of property rights);
- Policy neglect (as in the case of HIV/AIDS in poorer countries, which was allowed to linger and develop into a pandemic);

- Lack of knowledge and information (as in the case of chlorofluorocarbons, whose detrimental effect on the ozone shield was not recognized for a long time).

The publicness of a good in many instances reflects a policy choice. And so does globalness, which can be seen as a special dimension of publicness. From a worldwide perspective the creation of sovereign states has aspects of privateness, and reducing at-the-border or within-country policy barriers to free cross-border movements of goods, services, capital, and people as has happened during recent decades as a result of often difficult and lengthy intergovernmental negotiations entails reestablishing enhanced publicness.

The Link between Global Public Goods and Globalization Some global public goods such as the ozone layer or the moonlight are global-public by nature. Others would be most appropriately described as globalized national public goods. The globalization of hitherto essentially national public goods may occur for two main reasons. One is deliberate, intended harmonization of national public policy approaches and outcomes aimed at creating enhanced openness and interoperability of national physical and institutional infrastructure. Examples are the international civil aviation regime and the multilateral trade regime.

As openness progresses, national public policy domains become more closely interlocked. And as a result, a second, often unintended and also unaccounted-for process of national-public-good globalization sets in. Public effects or externalities now also move more easily across national borders communicable diseases, contagion effects of financial crises, crime and violence, including terrorist threats.

Thus a growing group of globalized national public goods is joining the natural global commons and contributing to the increasing policy attention that global public goods are commanding today, nationally and internationally.

The Provision Path of Global Public Goods In order to understand the provision path of global public goods, it is useful, first, to examine how public

goods are provided within the national context — and how this process has changed in recent decades.

Public goods and externalities are usually listed among the conditions associated with market failure and that may thus potentially justify state intervention. Many textbooks of public economics and public finance, therefore, refer to public goods also as state-provided goods. And indeed, some decades ago the state provided many goods directly and in full. In recent years, however, national borders have become increasingly open, the roles of markets and states have been rebalanced, and public-private partnering has increased. Today, public goods are often provided by multiple groups of actors — including governments, private business, civil society organizations, and private households and individuals.

Certainly, the state still exercises a certain amount of coercion (via regulation and taxation) to overcome the collective-action (free-riding) problems that often beset public goods precisely because they are public — available for all to consume once they exist. But its interventions are now geared less toward doing by itself what private actors won't do without adequate incentives, that is, direct intervention in the economy, and more toward creating incentives for other, nonstate actors to contribute to the provision of public goods.

If the provision of national public goods appears to be complex, the provision of global public goods is an even more complex process. Besides public and private actors it often also involves several — perhaps even, all — countries, and initiatives at the national as well as the international level.

The reason for the multilevel approach is that in many instances global public goods emerge as a result of a summation process: national public goods being provided in a concerted, harmonized fashion. For example, polio eradication requires immunizing the entire population in all countries. Of course, in order to achieve such concerted action across borders it is often important to have an international organization that oversees the implementation of the agreement to act. Thus global public goods may call for international cooperation *behind* national borders — within countries — and they may call for international cooperation *beyond* national borders — at the international level. However, the bulk of the international cooperation required in order to support global public goods in many instances occurs at the national level.

In cases such as promoting the safety of air travel through airport security checks, the provision level of the good (air safety) depends on the contribution of the weakest link in the chain of providers. In these cases, it may, from a certain provision level onward, be more efficient for the other providers to find ways and means of encouraging the weakest-link actors to upgrade their contribution rather than improving provision levels at home. In such weak-link situations, corrective action at home may need to be complemented by interventions abroad. Yet, even in these cases, the bulk of international cooperation would most likely still consist of interventions within nations.

An exception arises in cases where the global public good follows a "best shot" production path. The invention of a vaccine is a case in point. Once the new vaccine technology has been found and formulated, it exists. To reinvent it would be inefficient. Yet, in order to release the new vaccine technology and permit its use, the inventors may want to be reimbursed for the research and development cost they incurred. If the vaccine in question is one that would primarily be of interest to the poor, the international foreign-aid community might decide to share the reimbursement costs and organize an international collective-action initiative for this purpose. In this case, international-level efforts may perhaps outweigh related national-level, private-sector initiatives.

Global challenges often need global responses. But "global" does not necessarily mean interventions only at the supranational level. Rather, it primarily means that governments in many countries have to act, individually as well as collectively, and in many instances also together with private and civil society actors — in concert, so that all the various inputs fit together and the global public good actually emerges.

At the national level, the state can intervene in a more or less direct, more or less coercive or incen-

tivizing role to orchestrate the many contributions from the myriad of private and public actors who are often involved in the provision of public goods. However, the institution of the state has no equivalent internationally. Intergovernmental cooperation has to happen voluntarily. Hence a critical difference exists between the provision path of ("pure") national public goods and global public goods (including globalized national public goods). In international venues national governments are quasi-private actors: They tend to pursue national, particular interests.

One could thus expect that global public goods are severely underprovided. Indeed, several are, and it sometimes seems that the world is facing a lengthening list of global crises due to such underprovision a crisis of global warming, intermittent financial instability, international terrorism, or new and resurgent communicable diseases. But then one also has to bear in mind that many global public goods are quite well provided, facilitating large streams of cross-border economic activity, sharing of information and knowledge, or the universalization of human rights and other norms such as those that help facilitate a good business climate.

The reason why quite a large number of global public goods are adequately provided despite the fact that intergovernmental cooperation has to happen voluntarily lies in the fact that different actors have different preferences for various public goods. Where their preference for a particular good is sufficiently strong, actors may in effect not free ride but reveal their preference and pay for the provision of the good, irrespective of whether others may also enjoy it or not. Thus private business has often nudged governments into fostering the integration of markets. And civil society organizations have put pressure on governments to take action on global environment or gender concerns.

In the case of public goods generally, but certainly when it comes to global public goods, government failure or hesitation to act on these issues is often corrected by nonstate actors either by way of private and voluntary provision of these goods (as the growing incidence of industry self-regulation signals) or by way of political advocacy and lobbying.

Policy Implications and Challenges The growing importance of global public goods has changed public policymaking in fundamental ways. It confronts policymakers with new tasks to be met under new, changed conditions. As a result of greater economic openness, exclusive policymaking sovereignty is, in most countries, giving way to what might be called "responsive policymaking sovereignty." National policymaking increasingly takes into account not only national policy preferences but also external expectations (e.g., concerning rule of law or fiscal discipline) and global exigencies (e.g., the risk of global climate change).

As a result, it also appears that the role of the state is undergoing a fundamental change. The conventional Westphalian-type state that exercises exclusive policymaking sovereignty within its territorial boundaries is progressively giving way to the intermediary state, the hallmark of which is the blending of external and domestic policy expectations. Through this blending approach, states foster the globalization of national public goods and contribute what is expected of them in terms of delivery of national building blocks to global public goods.

The blending function of the intermediary state is evident not only from changes in budgetary allocations that echo global concerns (e.g., more resources devoted to fighting HIV/AIDS) but also from the use of new policy instruments geared to promoting, on the one hand, even further economic openness (e.g., through new investment laws), or on the other hand, new forms of national closure (e.g., through the introduction of "green" taxes to discourage border-transgressing pollution like greenhouse gas emissions).

Global public goods have changed the national toolkit of policymakers as well as conventional foreign policy. International cooperation beyond national borders is also changing from an essentially intergovernmental process to a multiactor process of public-private partnering. Similar to the change in the role of the state nationally, intergovernmental organizations are increasingly sharing the operational tasks of international cooperation with nonstate actors. Global markets are taking on tasks formerly

performed by intergovernmental organizations. Examples are the emerging international carbon markets or also those for commodity-related futures and options markets. A variety of global public-private partnerships   established as either nonprofit or for-profit organizations   are today producing inputs into the provision of global public goods and making contributions to global equity concerns such as the development of poorer countries.

The Current Stage of the Debate on Global Public Goods Global public goods and their provision are a reality of growing importance. However, at present, this reality is frequently viewed through "old" conceptual lenses. International cooperation "practitioners" often view global public goods provision as foreign aid. The reason is that foreign aid or development assistance is a well-entrenched institution with existing policy and resource-mobilization channels. Thus when a global challenge like avian flu control requires financial support for a weak-link country that may lack requisite means for decisive and prompt corrective action, the richer industrial countries   the conventional aid donors   often use aid resources to provide such support. They do so, although this support is perhaps not being provided for ethical or moral reasons alone (i.e., because the recipient country is poor) but primarily in the "donors'" own, national self-interest   for efficiency rather than equity reasons.

Misperceptions occur not only on the practical-political side of global public goods provision but also on the academic, theoretical side. For the most part, mainstream public goods theory, and more broadly, public economics and finance theory are still based on the assumption of a closed, single economy. Thus current mainstream theory of public economics/finance may not always capture well the total provision path of global public goods. Its national-level focus may make it difficult for researchers to recognize that these goods call for a *concerted* provision of relevant national public goods and for limited yet critically important *complementary* international-level inputs like international agreements that often provide the framework for national-level policy harmonization.

A full understanding of how the provision of global public goods works at present and how it could work more effectively, efficiently, and equitably would require a broad perspective   a comprehensive examination of the overall provision path of these goods. It would involve rethinking a number of aspects of the current mainstream theory of public finance/economics. Much of this rethinking has so far been undertaken in respect to particular issues and is available in specialized journals and debates.

The time is now ripe for undertaking a synthesis of both the theoretical innovations and the lessons learned from various practical-political innovations that have been undertaken in order to respond to global challenges under the pressure of today's changed policy realities. An important question to explore in this context might also be whether an adequate understanding and management of the global economy would call for the creation of a new subfield within the domain of public finance/economics   the field of "global public finance."

See also aid, international; globalization; health and globalization; international policy coordination; social policy in open economies

FURTHER READING
Barrett, Scott. 2007. *A Guide to Understanding Global Public Goods*. New York: Oxford University Press. Prepared as a background study for the International Task Force on Global Public Goods (see www.gpgtaskforce.org), this work aims to make the concept of global public goods accessible to policymakers and the general public, and identifies policy options for correcting the under provision of select global public goods that figure prominently on the international policy agenda.
Ferroni, Marco, and Ashoka Mody, eds. 2002. *International Public Goods: Incentives, Measurement, and Financing*. Washington, DC: Kluwer Academic and World Bank. Focuses on actions in the international arena to com plement domestic efforts in addressing global chal lenges and addresses strategic and practical measures, with a special emphasis on financing international co operation efforts in support of international   regional and global   public goods.

Kaul, Inge, and Pedro Conceição, eds. 2006. *The New Public Finance: Responding to Global Challenges.* New York: Oxford University Press. Recognizing that standard public economics and public finance theory has increasingly been adjusted to reflect the rebalancing of the roles of markets and states that has happened in recent decades but not yet to the greater openness of national borders, the authors of this volume explore how public finance has so far responded to globalization, notably the increased importance of global public goods and development concerns.

Kaul, Inge, Pedro Conceição, Katell Le Goulven, and Ronald U. Mendoza., eds. 2003. *Providing Global Public Goods: Managing Globalization.* New York: Oxford University Press. Building on the research and policy debates stimulated by the aforementioned book, this volume advances the concept of global public goods and deepens the discussion on the provision of these goods, covering both the negotiations or policymaking side of international cooperation as well as the operational side of cross border collective action.

Kaul, Inge, Isabelle Grunberg, and Marc A. Stern, eds. 1999. *Global Public Goods: International Cooperation in the 21st Century.* New York: Oxford University Press. Introduces the concept of global public goods, explores its application to a variety of global concerns, and discusses the implications of the growing importance of global public goods for policymaking at the national and international levels.

Sandler, Todd. 2004. *Global Collective Action*. Cambridge: Cambridge University Press. Identifies the factors that promote or inhibit successful collective action at international levels for the growing set of global challenges associated with globalization. Because many of these concerns involve strategic interactions dependent on one's own and others' actions, the author examines these interactions in a simple but rigorous game theoretic framework.

Sweden, Ministry for Foreign Affairs. 2001. "Financing and Providing Global Public Goods: Expectations and Prospects." Prepared by Keith Bezanson and Francisco Sagasti on behalf of the Sussex University Institute of Development Studies. Stockholm. A study of the idea of an effective and well funded multilateral system for development. It recognizes that under conditions of globalization, development, including poverty reduction, and the provision of global public goods are two distinct yet closely linked endeavors.

www.gpgnet.net. The interested reader will find a comprehensive list of recently published contributions to the topic of global public goods, including academic studies, policy statements, and journalistic writings on the Knowledge Portal page. The Knowledge Portal classifies entries according to issue areas (e.g., climate stability, communicable disease control, or financial stability) and according to cross cutting themes (like the concept of public goods and global public goods or the politics and financing of global public goods provision).

INGE KAUL

■ globalization

Globalization refers to a multidimensional process whereby markets, firms, production, and national financial systems are integrated on a global scale. These processes can range across a wide variety of affairs, but all point to the internationalization of everyday life. While we could speak of purely political or cultural aspects of this "shrinking of the world," the economic dimensions of globalization attract the greatest attention. Economic globalization centers on the material mechanisms connecting different countries, which then have political, cultural, and social consequences within and between states.

Some specific definitions of globalization take an overarching view of the economic ties comprising the phenomenon. For instance, the Organisation for Economic Co-operation and Development defines globalization as a process "by which markets and production in different countries become increasingly interdependent due to dynamics of trade in goods and services and then flows of capital and technology" (Stocker 1998). Others try to focus on more limited aspects of the economic contact across borders. Some, such as Richard Rosecrance (1999), prioritize the ease and volume of international flows of the factors of production themselves. For simplicity's sake, we can limit the discussion to three factors of production: land, labor, and capital. Land

obviously is never mobile from one state to another, but the ability to move labor and capital over borders has fluctuated over time. Louis Pauly (1997) has observed that the opening up of financial markets since the early 1980s has focused scholarly attention on the fluidity of capital across borders. For many who study globalization, the integration of national financial systems, symbolized by the ease and speed at which capital can flow internationally, is the critical element of globalization.

Others emphasize the volume of goods in trade. Ronald Jones (2000), an economist, describes globalization in terms of rising interdependence thus emphasizing the nature of the goods involved in trade, not merely the volume. This sort of interdependence involves a high degree of specialization and may also entail a fairly high level of integration of production across borders. While we can imagine ways in which factor flows involve single episodes of capital or labor flowing from one location to another (leaving each state different from before, but without a lasting relationship), the emphasis on trade suggests an evolving, intertwining connection. Since trade involves altering the mix of goods produced locally, and production in turn shapes how people work, where they live, and what they can consume, trade should be considered an essential part of globalization.

Given the many different definitions of globalization, scholars have employed a wide variety of indicators to measure this phenomenon. Some people recommend using the value of exports plus imports as a percentage of gross national product (GNP) as a measure. Others note that the nature of the goods being traded must be changing, to reflect the interdependent nature of the trade itself; this would require a measure that captures the changes in types of goods exported and imported over time. The pattern of flows of goods or factors is also important for distinguishing globalization from other phenomena such as bilateral or regional integration. For those interested more in factor flows, we would turn to measures of economic migration (such as percentage of the population born abroad) or of capital flows (again perhaps contrasted against a measure of

domestic economic activity such as domestic savings or GNP).

Earlier Examples of Economic Globalization If we look at the evidence generated by any one of these indicators, we would discover that globalization is not a new phenomenon. The degree of the globalization of the world economy has fluctuated over time. For instance, during the decades between 1870 and 1914, national economies became much more tightly integrated than they had ever been before.

Economic globalization gained momentum in the late 19th century for several reasons. Technological advances in the early and middle part of the century reduced the cost of long-distance transportation. In particular, the harnessing of steam power made it possible to ship bulky commodities very far, very rapidly, at greatly reduced costs (O'Rourke 1997). Railroads penetrated deeply into the continents of North and South America, as well as Europe and Australia. Steamships shortened the time it took to cross oceans and also freed the ships to travel in directions other than the prevailing winds. The other important technological advance of this period involved communications. The development of the telegraph allowed people the first opportunity to send messages almost instantaneously over great distances. The ability to send information rapidly critically altered several aspects of the world economy. For instance, rapid communications helped individuals and firms select when and where to ship their goods. More important, better communications made it possible to integrate international production. Before the mid-19th century, the slow speed of communications made it difficult for producers to react to changes in distant markets or to seize new opportunities as they emerged.

Of course, policy decisions also determined why globalization accelerated in these years. The spread of liberal ideology in the middle of the 19th century led many states to lower their tariffs and to join the gold standard. By lowering their tariffs, they made it possible for firms and individuals to enter into international trade. These changes then triggered trade adjustment, as market forces pushed countries toward specializing in the types of goods they had a

comparative advantage in producing. The gold standard had attributes that also supported economic globalization. By stabilizing exchange rates, the gold standard reduced uncertainties in international transactions. It also committed states to open flows of capital. Considering the high levels of international migration that also occurred in these decades, it is no surprise that many people argue the world's economy was more globalized in this era than in any other capital and labor were both highly mobile across borders, as were all types of goods (i.e., both manufactured and agricultural products).

Economic globalization had risen and fallen in even earlier periods. In the 17th and 18th centuries, the expansion of the empires of European states caused the integration of their colonies' economies with those of the imperializing states. This economic integration was driven by mercantilism, however. Therefore, although parts of the world economy were linked together over vast distances, the type of integration was much more of the "hub and spoke" model than truly globalizing. In other words, we might observe the development of long-distance trade and the introduction of new commodities into international exchange (such as tobacco or chocolate), but the flow of goods often occurred within the political jurisdiction of a single authority rather than across the boundaries of states. Also, transportation and communication remained relatively slow.

Technological advances often triggered these earlier examples when globalization accelerated. The integration of the world's economy in the 16th and 17th centuries depended on earlier advances in navigation, the ability to construct more seaworthy ships, and map-making. These advances made it possible to introduce not only long-distance trade (as between Europe and Asia) in expensive items such as spices or precious metals but also to include bulky commodities such as timber or grain across shorter distances. While improvements in technology may permit globalization to proceed, by themselves improvements in technology cannot cause globalization. Some actor must seize the technology and utilize it; moreover, state policies must allow globalization to continue. Thus while technological ad-

vances create new opportunities for globalization, the impact of those opportunities is not felt evenly across the world.

Learning from These Earlier Eras of Globalization There are some important lessons to be learned from these earlier experiences with globalization. In the late 19th century, for instance, economic globalization triggered a greater integration of national economies, but various countries responded to this integration in quite different ways. Some countries grasped the economic opportunities present and used them to improve their economic welfare. In others, globalization sparked a political backlash.

As noted earlier, economic globalization initiates change. Some of these changes may appear fairly positive, but most involve reorganizing economic activity and consequently redistributing wealth within societies. Liberalizing trade is a perfect example. Liberalizing trade creates more wealth for society as a whole, but it also redistributes wealth within an economy. In the late 19th century, globalizing trade relations triggered both extreme efforts at adjusting production to exploit the opportunities present and also extreme efforts to shield national economies and societies from the pressures to transform themselves. Some states embraced liberalized trade, changed the goods they produced (and thus adopted new production techniques as well), and consequently reorganized the way people worked and lived. Countries such as Britain, Belgium, Sweden, and Denmark made great strides economically by industrializing and shifting to smaller-scale (but more capital-intensive) agricultural production. These changes were intimately tied not only to the creation and redistribution of additional wealth, but also to urbanization and dramatic changes in land-holding patterns in rural areas. These transformations in turn created new political dynamics in these countries. The economic improvements that came with globalization can be linked with additional pressures for democratization and the transformation of political institutions.

On the other hand, some states were politically dominated by groups that feared globalization. In

places such as Germany, the landed elite used their political advantages to defend their social and economic status. This meant separating the national economy from global ties. By blocking trade adjustment, the landed elite could continue to sell the goods they produced at prices above world market levels — but this undoubtedly proved a drag on other parts of the economy. The elites refused to share political power, since that would have led to a decline in their economic and social status as well. The elites' preference for protection led them to oppose democratization. Before leaping to the conclusion that protection and authoritarianism go hand in hand, however, one must consider other cases from the late 19th century. In France, the Third Republic's electoral rules gave great voice to the rural areas. Land reforms in the wake of the French Revolution had taken away land from large landholders and given it to peasants in small plots. Economic globalization proved more of a threat than an opportunity to these small farmers. Adjusting their production proved difficult, since they lacked the capital or technical skills to adopt new techniques. Their inability to compete in the global economy, coupled with their strong representation in the democratic institutions, led to overwhelming support for protectionism. In short, drawing on the experiences of a wide range of countries in the late-19th-century era of globalization, we can conclude that there is no simple relationship between the economic and political consequences of globalization. Depending on the economic and political context, globalization's impact varies.

Studying the Current Period of Globalization
The descriptions given earlier suggest that technological improvements may be a necessary (though not sufficient) cause for globalization to proceed. Different disciplines emphasize different dimensions of globalization but also pose quite varied questions about why it occurs or what its consequences are. This diversity can be observed by examining how separate disciplines address the issue.

For economists, the reasons for globalization often go unquestioned. Expansion of market relations makes perfect sense within standard economic frameworks. If markets are tools for generating greater utility for participants, then individuals and firms will seek to participate in ever wider markets. Economists may ask more questions about the consequences of this market integration — which is also their way to understand when and where opposition to globalizing forces arises. Depending on the aspects of economic globalization one selects, economists have developed models depicting how the increase in economic ties will affect domestic actors. While an expansion in the size and scale of the market generates growth and productivity gains, specialization triggers changes that redistribute wealth in important ways. Since most of these changes create both winners and losers, economists may therefore recommend that states participate in globalization and then adopt policies that compensate those who do not enjoy the benefits generated. Evidence also suggests that states that open themselves up to the world economy adopt higher levels of social policies to shield segments of the domestic economy from the buffeting caused by global economic forces (Rodrik 1997). On the whole, economists endorse globalization for the material gains it makes possible, although they recognize that it can also bring instability.

As the brief descriptions of previous examples when globalization accelerated make clear, globalization both is the result of policy choices shaped by politics and entails political consequences. Political scientists are interested in not only the consequences of globalization, but also why globalization has not proceeded in a clear, smooth trajectory over time. To understand why states choose to participate in economic globalization, political scientists begin with an economic model describing the likely economic gains and their distribution within the state, then use that information to inform their understanding of how interests align in political competition. Using models developed in political science to understand how institutions shape this competition over policy, political scientists can then explain why some states pursue the economic opportunities globalization offers (willingly making the domestic adjustments necessary) while other states forgo those economic opportunities.

Political scientists also question why globalization has proceeded in waves. While technological change may provide some insights into this pattern, political scientists also point to possible changes in the international political system. Markets require underpinnings to function effectively one need only think of the importance of an accepted medium of exchange as an example. Most national markets function well because these infrastructural underpinnings are provided by a legitimate political authority. Yet the international system lacks a similar political authority. Instead it is composed of numerous actors claiming to be sovereign. Thus political scientists would emphasize the importance of powerful or leading states (or perhaps of actors within powerful states) who could provide and manage international institutions that make economic globalization possible. Although the technological changes of the mid-19th century triggered a great expansion in the flow of goods and factors of production across national boundaries in the pre World War I period, that war disrupted many of those patterns. In the 1920s and 1930s, national policies undermined efforts to re-create a global economy but many parts of the infrastructural underpinnings were also lacking. The United States and its allies made a conscious effort to create institutions fostering economic globalization after World War II.

It should come as no surprise that political scientists and economists are interested in globalization's impact on individual state's policies. On the one hand, many fear that integration into larger markets increases the pressures to be competitive. Globalization has therefore been viewed as a threat to states' freedom of choice on many different fronts. If one assumes that intervention in markets leads to distortions that reduce the competitiveness of domestic firms or individuals, or if one believes that governments are largely parasitic (i.e., not necessary), then one might believe in the "race to the bottom" thesis: the claim that economic globalization pressures states to deregulate markets more and more. However, there is little evidence of such a general trend. Instead, data suggest that states have responded to economic globalization in a variety of ways, with some deregulating markets, but others introducing more government activity or services (Burgoon 2001). If the government can provide services that add to domestic actors' competitiveness, then globalization may trigger greater involvement in the economy, not less.

On the other hand, the pressures to compete also suggest ways in which government policies aimed at noneconomic goals may be adversely affected by globalization (Rudra 2002). Governments have increasingly sought to borrow funds to pay for their programs. Integration of national financial systems leads governments to borrow from foreign sources. Since their expenditures may target goals such as social or economic equality, without the aim of increased economic competitiveness, these states must be able to bear the weight of the debts they carry. Such pressures may be especially problematic for economically developing countries, since they may also wish to resist the global market's demands to specialize in nonindustrial production.

If we concentrate on international flows of the factors of production, the current period of globalization is striking for the speed and amount of short-term capital flows. Although high levels of capital were invested abroad in the late 19th century, this investment was largely long term (Williamson 1996). In the current period, the value of short-term speculative international investments now outweighs the value of goods in trade. Capital flows of this nature may deliver fewer benefits than the risks they also carry, especially for countries with few hard currency reserves. Therefore, many scholars have begun to question whether capital controls might be valuable suggesting that we have reached the limits of globalization. Labor, on the other hand, does not flow very freely and in the post September 11 world, flows of labor are not likely to be freed up anytime soon.

Globalization and Identity Globalization can describe more than economic relations, of course. Since other dimensions of globalization can refer to communications or media, globalization is often also related to cultural changes. These aspects may trigger political responses on their own or interact with the

politics sparked by economic globalization. Cultural issues may arise when we examine the nature of the goods crossing borders during globalization. If the goods include entertainment products (such as books, music, and films), these economic interactions can unleash strong forces on society.

This imagery was captured beautifully by the journalist Thomas Friedman in *The Lexus and the Olive Tree* (1999). Friedman's two images signify the pros and cons of globalization. The Lexus represents the greater wealth globalization makes possible. The olive tree represents culture the way domestic actors are rooted. In Friedman's view, domestic actors want both: they desire continuation of their culture (with minimal outside influence) but also greater wealth. Friedman's symbols capture very well the test globalization presents to societies: how can a state ensure that its engagement in the global economic system brings more benefits than costs?

Challenges Ahead in the Study of Globalization Globalization presents several challenges for social scientists, but the greatest is undoubtedly how to model the impact of the many dimensions of globalization together. This requires multidimensional modeling, since trade and factor flows can each have different impacts. On top of these effects, we would want to add different domestic policies that might moderate the impact of increased immigration, international capital flows, or trade adjustment. A broader model would also need to integrate the politics of identity triggered by globalization. Only when we bring together different theoretical approaches from international and domestic politics, and from micro- and macroeconomics, will we have a better grasp on the broad phenomenon we call globalization.

See also anti-globalization; capital mobility; comparative advantage; exchange rates and foreign direct investment; foreign direct investment (FDI); migration, international; nontraded goods; social policy in open economies; trade in services

FURTHER READING

Burgoon, Brian. 2001. "Globalization and Welfare Compensation: Disentangling the Ties that Bind." *International Organization* 55 (3): 509 51. Examines the way domestic policies offset some of the risks economic openness creates.

Frieden, Jeffry. 1991. "Invested Interests: The Politics of National Economic Policies in a World of Global Finance." *International Organization* 45 (4): 425 51. A clear (and often cited) examination of the domestic politics on international monetary policies.

Friedman, Thomas. 1999. *The Lexus and the Olive Tree*. New York: Farrar, Straus and Giroux. An easy to read, nonacademic discussion of the way economic globalization interacts with the politics of identity.

Garrett, Geoffrey. 1995. "Capital Mobility, Trade, and the Domestic Politics of Economic Policy." *International Organization* 49 (4): 627 55. Provides an overview of the arguments and evidence regarding the way economic globalization reorganizes domestic interests, and the consequences that has on politics.

Garrett, Geoffrey, and Peter Lange. 1996. "Internationalization, Institutions, and Political Change." In *Internationalization and Domestic Politics*, edited by Helen Milner and Robert Keohane. Cambridge: Cambridge University Press, 48 75. Maps out various ways to model the connections between the domestic and international political economies.

Hirst, Paul, and Grahame Thompson. 1999. *Globalization in Question: The International Economy and the Possibilities of Governance*. 2d ed. Cambridge: Polity Press. A classic textbook on globalization and global governance.

Jones, Ronald. 2000. "Private Interests and Government Policy in a Global World." Tinbergen Institute Discussion Paper TI 2000 051/2. Amsterdam, the Netherlands: Tinbergen Institute. An accessible economic analysis of economic openness.

O'Rourke, Kevin. 1997. "The European Grain Invasion, 1870 1913." *Journal of Economic History* 57 (4): 775 801. Concise example of how economic globalization affected the production and consumption of a single commodity in an earlier period of globalization, and the economic and political consequences that followed.

Pauly, Louis. 1997. *Who Elected the Bankers?* Ithaca, NY: Cornell University Press. Questions the responsibility and accountability placed on those who make international financial policy.

Rodrik, Dani. 1997. *Has Globalization Gone Too Far?* Washington, DC: Institute for International Economics. An economist's attempt to assess the impact of globalization.

Rosecrance, Richard. 1999. *The Rise of the Virtual State.* New York: Basic Books. Creatively explores the way technology is changing both the economic connections between states and also the way states relate to each other politically.

Rudra, Nita. 2002. "Globalization and the Decline of the Welfare State in Less Developed Countries." *International Organization* 56 (2): 411 45. Presents evidence concerning the relationship between international openness and the capacity of economically developing countries to provide social welfare policies.

Stocker, Herbert. 1998. "Globalization and International Convergence in Incomes." In *Economic Effects of Globalization*, edited by John ren Chen. Aldershot, UK: Ashgate, 97 115. An interesting contribution to an edited volume, both for its discussion of the definition of globalization and for the way in which it measures the results of liberalizing international economic policies.

Williamson, Jeffrey. 1996. "Globalization and Inequality Then and Now: The Late Nineteenth and Late Twentieth Centuries Compared." NBER Working Paper No. 5491. Cambridge, MA: National Bureau of Economic Research. A masterful overview of the domestic economic consequences of globalization in several cases drawn from the late 19th century.

MARK R. BRAWLEY

■ gold standard, international

The international gold standard was a grand experiment in decentralized monetary unification. From approximately 1880 to the start of World War I, many countries in Europe, North America, and the Southern Hemisphere spontaneously joined the gold standard   a monetary system in which the value of money is defined in terms of a given quantity of gold   during a time of free capital and labor mobility and low levels of protectionism. These countries were diverse in income levels and stages of economic development, but shared the belief that the conservative handling of money would ensure price stability, and that central banks should enjoy a high degree of independence from government.

The international gold standard was a decentralized process but had its own rules (McKinnon 1993). To begin with, paper money had to be convertible into a fixed weight of gold at an official price. Convertibility could be suspended under exceptional circumstances, such as wars or banking and financial crises. After a temporary period, however, convertibility would have to be reestablished at the original parity. Second, the expansion of the national money stock was limited by the stock of monetary gold. Third, gold and capital were free to move from country to country. Finally, monetary authorities were not to hamper the economic adjustment that was implied by gold flows; on the contrary, they were to reinforce this adjustment with appropriate policy actions.

Bilateral exchange rate parities were determined by the relative money prices of gold. As an example, in 1879 the official price of an ounce of gold was £4.252 in Britain and $20.67 in the United States. Parity between the dollar and the pound was, therefore, $20.67/£4.252 or $4.856 = £1.0. The exchange rate fluctuated around parity but was constrained by a band, of which the upper and lower edges   called the gold points   were defined by the transaction costs of shipping gold from one country to another. In practice, the bilateral bands of fluctuations were rather narrow. For the dollar-pound exchange market, the upper edge was estimated at $4.890 and the lower edge at $4.827. If the actual exchange rate were to move beyond $4.890, gold would flow from the United States to the United Kingdom; if the exchange rate were to fall below $4.827, gold would move in the opposite direction. The band of exchange rate fluctuations was and is a typical feature of fixed exchange rates.

Basic Principles The gold standard rested on a decentralized organization in which participants retained their own monetary authorities, their own fiscal autonomy, and their own political sovereignty; in essence, participating countries enjoyed monetary integration without fiscal and political integration. If properly functioning, the system ensured a common

price level, the world price level, which was determined by the interaction of the global supply of and the global demand for money. The global supply of money was proportional to the stock of monetary gold; the global demand for money was the weighted sum of the national demands for money. The world price level would fall when real income rose relative to new flows of monetary gold. The latter would respond positively to the price of gold and productivity improvements in gold mining, and negatively to the world price level. As the price level rose, the stock of gold moved from monetary to industrial use. The decline in the stock of monetary gold reduced, in turn, the price level. For the world price level to be stable, the stock of monetary gold, inclusive of new flows, had to act as a buffer against the forces affecting the money market. But the stock of monetary gold, in fact, behaved too erratically to play this role (Cooper 1982, table 3), and prices, as we will see later, were not stable.

A country with a price level misaligned with respect to the world price level would face an adjustment process. The details of this process, known as the price-specie flow mechanism, were worked out by philosopher and economist David Hume in 1752. For example, a country experiencing a higher price level than the world price level would experience a balance-of-payments deficit. This deficit, in turn, would force a gold outflow and a proportional decline of the domestic monetary base. The money stock would decline in sympathy with the monetary base. With the economy running at capacity, the decline in the money stock would lower the domestic price level. The gold outflow in the deficit country would be directed toward countries with balance-of-payments surpluses and a lower price level than the world price level. In the surplus countries gold inflows would raise the monetary base and generate a process opposite to that in the deficit countries.

Under the assumption of fully utilized resources and price and wage flexibility, deficit and surplus countries would restore price level equilibrium symmetrically. To speed up adjustments, central banks were supposed to reinforce gold flows with appropriate policy actions. A gold outflow would have to be followed with a rise in the discount rate (the interest rate a central bank charges to commercial banks and other financial intermediaries for loans of reserve funds), which would attract capital inflows and contract at the same time the domestic component of the monetary base and bank credit; the opposite for a gold inflow. In essence, the rules of the game prescribed a positive correlation between the foreign and the domestic components of the monetary base.

Basic Facts The practice of the gold standard differed somewhat from the theory. The system actually evolved into a gold reserve standard, with "senior" and "junior" currencies and financial centers and a periphery. The British pound was the "senior" currency and London the biggest and most sophisticated money and financial center in the world. The French franc and the German mark were junior currencies; Paris and Berlin, junior centers. All other countries were periphery, although with different degrees of closeness to the senior and junior centers. Foreign exchange reserves   that is, short-lived assets denominated in the currencies of the centers   grew more rapidly than gold reserves. The amount of international liquidity held in the reserve centers by far exceeded the total gold reserves held by these countries, in particular British official gold. The senior center and the two junior centers were net capital exporters, reassuring reserve holders that the centers were financially solid. All in all, there was a clear tendency for the gold standard to evolve into a gold reserve standard, economizing on gold.

The adjustment mechanism was not as symmetric as the theory claimed. Instead, a hierarchical structure was in place: the same action imparted different effects on the system depending on where it was initiated. The senior center had the largest impact; the junior centers had less and the periphery even less. Hierarchy, or asymmetry, did not imply that the actor located at a higher level of the pyramid was fully insulated from the effects of the actions taken by actors located at a lower level of the pyramid. Capital flows were more sensitive to changes in British interest rates than to interest rates in other countries. Increases in the Bank of England bank rate (the equivalent of

the discount rate) tended to be followed by increases in the discount rates of other countries. Furthermore, if other countries were to match the British bank rate increase, capital would still flow to London.

To stem reserve outflows, and thus to prevent a convertibility crisis, other countries had to more than match the bank rate increase (Lindert 1969). Britain also reacted to interest rate changes initiated elsewhere, especially in the two junior centers. Some countries Argentina, Bulgaria, Chile, Italy, Mexico, and Portugal were so low in the hierarchy that were unable to compete in retaining reserves during rounds of bank rate increases and they had either to devalue their currencies or leave the standard altogether. The pressure on the periphery countries was especially intense during times of gold shortage, causing deflationary pressure and relatively high real interest rates.

The rule of reinforcing gold flows with appropriate changes in the discount rate was widely ignored. The typical pattern was an inverse correlation between the foreign and domestic component of the monetary base, evidence that central banks were sterilizing (compensating the effect of gold flows so as to leave the monetary base unaffected) gold flows to pursue domestic objectives (Nurkse 1944; Bloomfield 1959). The U.S. Federal Reserve system, as an example, sterilized gold movements for much of the 1920s. Even the Bank of England, at the center of the system, violated the rule, although less than other central banks. The sterilization of gold inflows by surplus countries, coupled with the pressure on deficit countries to raise interest rates and contract output, contributed to a deflationary bias in the system; see Keynes (1923, chap. 4).

The main achievement of the international gold standard was to stabilize the long-run inflation rate. For example, the average annual inflation rate in the United Kingdom from 1870 to 1913, measured with reference to wholesale prices, was 0.7 percent; in the United States, from 1879 to 1913, 0.1 percent (Bordo 1981, table 1). The money supply in the system was constrained by gold discoveries. Individual countries, on the whole, were committed to parity and were willing to undertake anti-infla-

tionary policies to restore that parity if they departed temporarily from it. The strategy of sticking to a course of action over time promised payoffs in terms of lower interest rates. In fact, countries that were stricter in the application of the standard enjoyed lower interest rates than countries with a sporadic commitment to convertibility.

The ex-post long-run stability of the price level did not mean, however, that price level uncertainty was low. The coefficient of variation of the inflation rate calculated as the ratio of the standard deviation of the inflation rate to its mean was approximately 3 times higher during the gold standard than in the interwar period and 12 to 14 times higher than post World War II.

These short-run variations must have created considerable difficulty in the private sector in distinguishing price level changes from relative price movements. Furthermore, high short-run variability of the price level was made worse by a high degree of unpredictability of the adjustment period required for the price level to return to equilibrium level (Phinney 1932 33). The swings of the price level lasted up to 50 years. While it is true that prices in 1910 were practically the same as prices in 1850, people living in 1870 or in 1895 would have faced considerable uncertainty in predicting when prices would have returned to the 1850 level. One arrives at a similar conclusion by looking at the wholesale price indexes of four countries: the United Kingdom, Germany, France, and the United States from 1816 to 1913 (see Cooper 1982, table 2). Prices fell and rose dramatically over extremely long swings. For example, in the United States prices fell 45 percent from 1816 to 1849, rose 67 percent from 1849 to 1873, fell 53 percent from 1873 to 1896, and rose 56 percent from 1896 to 1913. The swings were somewhat smaller in the United Kingdom, Germany, and France.

The Gold Standard in the Interwar Period World War I brought an end to the international gold standard. The Brussels Conference of 1920 and the Genoa Conference of 1922 tried to restore fiscal discipline, central bank independence, and the gold reserve standard. Despite the loss of its preeminent

position as the leading country of the international monetary system, the United Kingdom was a strong voice in the two conferences. On the other hand, the United States, anointed to replace the United Kingdom, decided not to participate at Genoa for reasons related in part to disputes over war debts and reparations (Eichengreen 1992, 11).

France strongly opposed the establishment of a gold reserve standard at the Genoa conference, seeing in this an attempt by Britain to reassert the dominance of the pound and London. What followed were a series of ad-hoc decisions without a common design. Germany returned to the gold standard in 1924 in the aftermath of its disastrous hyperinflation. Britain returned to the standard in 1925, and France in 1928. Prewar parities were set in Germany and Britain. The rules of the game dictated that, should a country leave the standard, it should reenter it at the old parity. This commitment was supposed to enforce credibility and prevent destabilizing speculation. In *A Tract on Monetary Reform*, John Maynard Keynes (1923) predicted that the return of Britain to prewar parity would be extremely deflationary: it overvalued the pound and created a deficit in the current account. Contrary to the hopes of the prewar parity advocates, the necessary downward price adjustment, especially of money wages, did not materialize; the adjustment fell primarily on production and employment.

France's new parity undervalued the French franc, causing a surplus in the current account. Furthermore, widespread speculation of a future appreciation of the French franc was responsible for large capital inflows. Consequently, international reserves exploded and the Banque de France sterilized the inflows of reserves with sales of domestic assets to prevent an appreciation of the franc. Saddled with large nongold reserves, the French were reluctant to give Britain the privilege of financing current account deficits with pound-denominated debt. The French opted to redeem reserves into gold, causing deflation in the "center" and easing the appreciation pressure on the franc. In terms of the implied rules of the game, France was no longer willing to accept the pre World War I hierarchical structure.

Other countries as well were unwilling to cooperate. National policies became less internationally oriented as the Great Depression of the 1930s affected much of the industrial world. National political systems had become increasingly reluctant to tolerate the large swings in unemployment and cyclical movements that accompanied the commitment to parity, and the international gold standard eventually lost its appeal.

See also balance of payments; banking crisis; Bretton Woods system; commodity-price pegging; convertibility; currency crisis; foreign exchange intervention; international financial architecture; international liquidity; international reserves; money supply; quantity theory of money; sterilization; Triffin dilemma

FURTHER READING

Bloomfield, Arthur I. 1959. *Monetary Policy under the International Gold Standard*. New York: Federal Reserve Bank of New York. This study, together with that by Ragnar Nurkse, shows that during the gold standard central banks were pursuing domestic objectives in addition to convertibility.

Bordo, Michael D. 1981. "The Classical Gold Standard: Some Lessons for Today." *Review of the St. Louis Federal Reserve Bank* 63: 2 17. This article, written at the end of the worldwide inflation surge of the 1960s and the 1970s, reviews the record of the gold standard in providing stable prices and economic stability.

Cooper, Richard. 1982. "The Gold Standard: Historical Facts and Future Prospects." *Brookings Papers on Economic Activity* 1: 1 56. Reviews the economic record of the international gold standard especially with regard to the long swings in the price levels.

Eichengreen, Barry. 1992. *Golden Fetters. The Gold Standard and the Great Depression, 1919 1939*. New York: Oxford University Press. Explores the connection between the gold standard and the global economic crisis of the 1930s.

Hume, David. 1898 (1752). "On the Balance of Trade." In *Essays, Moral, Political and Literary*. London: Longmans, Green. This is the classic article on the so called price specie flow mechanism.

Keynes, John Maynard. 1923. *A Tract on Monetary Reform*. London: Macmillan. This is Keynes's famous study on the deflationary bias of the international gold standard.

Lindert, Peter H. 1969. "Key Currencies and Gold, 1900 1913." Princeton Studies in International Finance No. 24. Princeton, NJ: Princeton University. Presents evi dence that the international gold standard had a hierar chical structure, with a center and a periphery.

McKinnon, Ronald I. 1993. "The Rules of the Game: In ternational Money in Historical Perspective." *Journal of Economic Literature* 31: 1 44. Focuses on the rules of the game of the international gold standard and other monetary regimes.

Nurkse, Ragnar. 1944. *International Currency Experience*. Princeton, NJ: League of Nations. See description under Bloomfield, above.

Phinney, J. T. 1932 33. "Gold Production and the Price Level: The Cassel Three Percent Estimate." *Quarterly Journal of Economics*, 47: 647 79. An early article showing that the price level during the gold standard was subject to long and uncertain swings.

MICHELE FRATIANNI

■ government procurement

Government procurement refers to state purchases of goods and services and to the leasing of capital equipment. The absence of any binding multilateral rules on government procurement practices is re garded as one of the most significant lacunae of the World Trade Organization (WTO) and its prede cessor, the General Agreement on Tariffs and Trade (GATT). A plurilateral agreement on government procurement exists, but its membership is confined principally to industrialized countries. Many free trade agreements (FTAs) contain provisions on government purchases, and a nonbinding United Nations Model Law on Procurement of Goods, Construction, and Services adds to the uneven patchwork of international disciplines on public procurement matters.

Typically the objectives of government procure ment policies are not just economic, such as obtaining "value for money." Often procurement policies seek to promote the well-being of certain societal groups, firms, regions, and industries. For example, small- and medium-size firms are the beneficiaries of such fa voritism in many jurisdictions, including in the Uni ted States. State purchasers have plenty of means to attain these ends. Procurement procedures can be broken down into three distinct stages: namely, ten dering procedures, evaluation of bids, and the notifi cation of contract awards and associated review pro cedures. It is important to bear in mind that discrimination in a government procurement proce dure can be both de jure (and, therefore, potentially transparent) and de facto (especially where state offi cials are given considerable discretion).

There is considerable uncertainty over the total value of government procurement of goods and services. The last attempt to compile a comparable cross-country dataset on state procurement expen ditures was based on 1998 data (OECD 2001). On average, state expenditures on goods and services accounted for 7.6 percent of the national income of Organisation for Economic Co-operation and De velopment (OECD) nations. The total amount of government spending on goods and services in the industrialized world amounted to $1.795 trillion. The comparable numbers for the developing coun tries included in this OECD study were 5.1 percent and $0.287 trillion, respectively. Information con cerning the ways in which governments discriminate against foreign bidders for state contracts is also hard to come by. However, the Trade Policy Reviews of the WTO do contain some useful information on the latter, yet no comprehensive database of procure ment-related barriers to international trade has ever been prepared. It is quite likely that paucity of data has held back research on this subject.

Data limitations have not deterred theoretical analyses of the effect of public procurement dis crimination on market outcomes and international trade. Although these analyses are almost always partial equilibrium in nature, in principle such dis crimination could have general equilibrium effects (see Evenett and Hoekman 2006 for a discussion of the latter). As far as procurement discrimination in markets is concerned, the modern literature was started by Baldwin (1970) and Baldwin and Rich ardson (1972). They demonstrated that in a perfectly competitive market a procurement ban on foreign

purchases by a state body will affect production levels, imports, and prices only if the quantity of goods demanded by that body exceeds total domestic supply at the moment the ban is imposed. If not, the procurement ban would merely induce a reshuffling of domestic and state purchasers between domestic and foreign suppliers. This finding, which other researchers have confirmed in alternative market structures, implies that not every discriminatory procurement policy reduces market access, so diminishing potential exporter interest in trade agreements that constrain such discrimination.

Many governments solicit bids for public contracts and consequently some analysts have examined the effect of discrimination against foreign bidders in an auction setting. McAfee and McMillan (1989) is the best known such analysis. They demonstrated that the introduction of a price preference, which inflates the foreign bids by a specified percentage before all of the qualifying bids are evaluated, can actually reduce the expected cost to the government of purchasing the good or service. The logic here is very similar to that of the terms-of-trade effects that can be induced by imposing a tariff on imports. Foreign firms partially absorb the price preference and so lower their (pre-preference-adjusted) bids. Domestic firms raise their bids but by less than the amount of the preference-adjusted foreign bids. If the distribution of foreign and domestic firms' costs is such that the probability of a foreign firm winning the contract is sufficiently high even after the price preference is imposed, then there are circumstances when the state's expected outlay falls. This result provides an efficiency-based rationale for the use of price preferences. Whether governments have the information necessary to optimally implement such price preferences is another matter. Moreover, retaliation by other trading partners may shift the calculus away from imposing price preferences.

From what is known about the international negotiations on public procurement matters at the WTO and the GATT it is not evident that these theoretical considerations have been given much weight. Instead the reluctance of some WTO members to develop multilateral disciplines on

public procurement matters has been attributed to concerns that the objectives of procurement policies (including industrial policy goals) may not be effectively met using alternative government policy instruments, to defensive steps taken by the recipients of the rents created by current procurement discrimination, to the suspicion that certain political parties or campaigns are directly or indirectly funded through public procurement policies, and to the concerns that multilateral disciplines would be too costly or burdensome to implement. In fact, in 2004 WTO members decided not to start negotiating new multilateral rules to enhance the transparency of government procurement practices. Consequently, the only prevailing WTO agreement on government procurement is a plurilateral one, the negotiation of which was concluded in 1994. This agreement has 38 signatories (27 of which are members of the European Union). Having said this, 19 WTO members are observers to this agreement, 8 of which are negotiating to join this plurilateral accord. In addition to requiring steps that promote the transparency of the signatories' public purchasing policies and practices and introducing faster procedures for losing firms to challenge state procurement decisions, this plurilateral accord also bans   for agreed sectors   the use of price preferences. The obligations of the 1994 plurilateral agreement are binding on its signatories, and disputes can be subject to the WTO Dispute Settlement Understanding. Many other forms of discrimination, including those that may prevent bids outright by foreign firms or raise their costs, were not banned in the 1994 Agreement suggesting that international disciplines in this area of government policymaking could be strengthened. The members of this plurilateral accord have sought its renegotiation, and in December 2006 a provisional agreement was reached (Anderson 2007).

It is important to appreciate that the WTO is not the only venue where international accords on public procurement matters have been developed. Binding obligations on these matters are frequently included in FTAs, in particular when an industrialized country is involved. In addition, the United Nations has developed a nonbinding Model Law on Procure-

ment, Goods, Construction, and Services. In the years to come researchers may want to examine whether the spread of FTAs with public procurement provisions alters WTO members' positions on the desirability of negotiating corresponding multilateral provisions. Moreover, attention could be given to the relative impact of these binding and nonbinding international initiatives on the reform of national procurement practices in both developing and industrialized countries.

See also nontariff measures; World Trade Organization

FURTHER READING

Anderson, Robert D. 2007. "Renewing the WTO Agree ment on Government Procurement: Progress to Date and Ongoing Negotiations." *Public Procurement Law Review* 16 (4): 255 73. An account of the agreement reached at the end of 2006 on revising the WTO's Government Procurement Agreement.

Baldwin, Robert E. 1970. *Nontariff Distortions of Interna tional Trade*. Washington, DC: The Brookings Institu tion Press. A classic introduction to nontariff barriers and of particular reference here to government pro curement practices.

Baldwin, Robert E., and J. David Richardson. 1972. "Government Purchasing Policies, Other NTBs, and the International Monetary Crisis." In *Obstacles to Trade in the Pacific Area: Proceedings of the Fourth Pacific Trade and Development Conference*, edited by H. E. English and Keith A. J. Hay. Ottawa: School of International Affairs, Carleton University, 211 303. Another early analytical treatment of the effects of government procurement practices on trade flows.

Evenett, Simon J., and Bernard M. Hoekman. 2006. "Introduction." In *The WTO and Government Pro curement*, edited by Simon J. Evenett and Bernard M. Hoekman. Cheltenham: Edward Elgar, vi xxii. In cludes all of the major legal, economic, and develop mental papers that discuss the relationship between trade policy, trade agreements, government procure ment policies, and international trade and economic performance more generally. The introduction to this volume contains a survey of the economic literature on the effects of procurement policies on national welfare and trade flows.

McAfee, R. Preston, and John McMillan. 1989. "Govern ment Procurement and International Trade." *Journal of International Economics* 26 (3/4): 291 308. Contains an analysis of the effects of procurement discrimination against foreign bidders in an optimal auction setting.

Organisation of Economic Co operation and Develop ment (OECD). 2001. "The Size of Government Pro curement Markets." *Journal of Budgeting* 1 (4). Avail able at http://www.oecd.org/document/63/0,3343,de 2649 201185 1845951 1 1 1 1,00.html. Contains statistics on the size of national procurement markets.

SIMON J. EVENETT

■ gravity models

Gravity models utilize the gravitational force concept as an analogy to explain the volume of trade, capital flows, and migration among the countries of the world. For example, gravity models establish a baseline for trade-flow volumes as determined by gross domestic product (GDP), population, and distance. The effect of policies on trade flows can then be assessed by adding the policy variables to the equation and estimating deviations from the baseline flows. In many instances, gravity models have sig nificant explanatory power, leading Deardorff (1998) to refer to them as a "fact of life."

Alternative Specifications Gravity models begin with Newton's law for the gravitational force (GF_{ij}) between two objects i and j. In equation form, this is expressed as:

$$GF_{ij} = \frac{M_i M_j}{D_{ij}} \quad i \neq j \qquad (1)$$

In this equation, the gravitational force is directly proportional to the masses of the objects (M_i and M_j) and indirectly proportional to the distance between them (D_{ij}).

Gravity models are estimated in terms of natural logarithms, denoted "ln." In this form, what is multiplied in equation 1 becomes added, and what is divided becomes subtracted, translating equation 1 into a linear equation:

$$\ln GF_{ij} = \ln M_i + \ln M_j - \ln D_{ij} \quad i \neq j \qquad (2)$$

Gravity models of international trade implement equation 2 by using trade flows or exports from county i to country j (E_{ij}) in place of gravitational force, with arbitrarily small numbers sometimes being used in place of any zero values. Distance is often measured using "great circle" calculations. The handling of mass in equation 2 takes place via four alternatives. In the first alternative with the most solid theoretical foundations, mass in equation 2 is associated with the gross domestic product (GDP) of the countries. In this case, equation 2 becomes:

$$\ln E_{ij} = \alpha + \beta_1 \ln GDP_i + \beta_2 \ln GDP_j + \beta_3 \ln D_{ij} \quad (3)$$

In general, the expected signs here are $\beta_1, \beta_2 > 0$. However, the economics of equation 3 can lead to the interpretation of GDP as income, and when applied to agricultural goods, Engels's Law allows for GDP in the destination country to have a negative influence on demand for imports. Hence it is also possible that $\beta_2 < 0$.

In the second alternative, mass in equation 2 is associated with *both GDP and* population (POP). In this case, equation 2 becomes:

$$\ln E_{ij} = \varphi + \gamma_1 \ln GDP_i + \gamma_2 \ln POP_i + \gamma_3 \ln GDP_j + \gamma_4 POP_j + \gamma_5 \ln D_{ij} \quad (4)$$

With regard to the expected signs on the population variables, these are typically interpreted in terms or market size and are therefore positive ($\gamma_2, \gamma_4 > 0$). That said, however, there is the possibility of import substitution effects as well as market size effects. If the import substitution effects dominate, the expected sign is $\gamma_4 < 0$.

In the third and fourth alternatives, mass in equation 2 is associated with GDP per capita and with *both* gross domestic product *and GDP* per capita, respectively. In these cases, equation 2 becomes one of the following:

$$\ln E_{ij} = \tau + \delta_1 \ln (GDP_i/POP_i) + \delta_2 \ln (GDP_j/POP_j) + \delta_3 \ln D_{ij} \quad (5)$$

$$\ln E_{ij} = \mu + v_1 \ln GDP_i + v_2 \ln (GDP_i/POP_i) + v_3 \ln GDP_j + v_4 (GDP_j/POP_j) + v_5 \ln D_{ij} \quad (6)$$

Since they involve the same variables, the parameters of equations 4, 5, and 6 are transformations on one another: $\gamma_1 = \delta_1 = v_1 + v_2$; $\gamma_2 = -\delta_1 = -v_2$; $\gamma_3 = \delta_2 = v_3 + v_4$; and $\gamma_4 = -\delta_4 = -v_4$.

Theoretical Considerations After being introduced by Tinbergen (1962), the gravity model was considered to be a useful physical analogy with fortunate empirical validity. Subsequently, however, connections have been made to key elements of trade theory. The standard assumption of the Heckscher-Ohlin model that prices of traded goods are the same in each country has proved to be faulty due to the presence of what trade economists call "border effects." Properly accounting for these border effects requires prices of traded goods to differ among the countries of the world. Gravity models have been interpreted in these terms.

Anderson (1979) was the first to do this, employing the product differentiation by country-of-origin assumption, commonly known as the "Armington assumption" (Armington 1969). By specifying demand in these terms, Anderson helped to explain the presence of income variables in the gravity model, as well as their multiplicative (or log linear) form. This approach was also adopted by Bergstrand (1985), who more thoroughly specified the supply side of economies. The result was the insight that prices in the form of GDP deflators might be an important additional variable to include in the gravity equations described above. Price effects have also been captured using real exchange rates (e.g., Brun et al. 2005).

The monopolistic competition model of New Trade Theory has been another approach to providing theoretical foundations to the gravity model (Helpman 1987; Bergstrand 1989). Here, the product differentiation by country-of-origin approach is replaced by product differentiation among producing firms, and the empirical success of the gravity model is considered to be supportive of the monopolistic competition explanation of intraindustry trade. However, Deardorff (1998) and Feenstra (2004) have cast doubt on this interpretation, noting the compatibility of the gravity equation with some forms of the Heckscher-Ohlin model and, consequently,

the need for empirical evidence to distinguish among potential theoretical bases: product differentiation by country of origin; product differentiation by firm; and particular forms of Heckscher-Ohlin-based comparative advantage. In each of these cases, the common denominator is complete specialization by countries in a particular good. Without this feature, bilateral trade tends to become indeterminate.

Alternatively, there are other approaches to gravity-based explanations of bilateral trade that do not depend on compete specialization. As emphasized by Haveman and Hummels (2004), this involves accounting for trade frictions in the form of distance-based shipping costs or other trade costs, as well as policy-based trade barriers. Distance costs can also be augmented to account for infrastructure, oil price, and trade composition, as in Brun et al. (2005). The two approaches (complete versus incomplete specialization) can be empirically distinguished by category of good, namely differentiated versus homogeneous, as in Feenstra, Markusen, and Rose (2001).

Assessment Due to its log-linear structure, the coefficients of the gravity model are in terms of elasticities or ratios of percentage changes. These "unitless" measures are comparable across countries and goods and give us direct measures of the responsiveness of trade flows to the trade potential variables of equations 3 6. For GDP and distance, estimated elasticities tend to be close to 1.0 in value. For distance, comparison across groups of countries gives a measure of the degree of integration in the world economy. In addition to these standard variables, the coefficients of policy variables help us to understand the impacts of the represented policies on trade flows. It is also possible to obtain estimates of border effects independently of distance and other variables, as well as to investigate some issues in economic geography, as in Redding and Venables (2004). Despite some ambiguity regarding its theoretical foundations, then, the gravity model is an important empirical tool to help us understand trade and other economic flows in the world economy.

See also applied general equilibrium modeling; Heckscher-Ohlin model; intraindustry trade; monopolistic competition; New Trade Theory; partial equilibrium modeling; revealed comparative advantage

FURTHER READING

Anderson, James E. 1979. "A Theoretical Foundation for the Gravity Equation." *American Economic Review* 69 (1): 106 16. A first attempt to provide theoretical foundations to the gravity model.

Armington, Paul. 1969. "A Theory of Demand for Products Distinguished by Place of Production." *IMF Staff Papers* 16 (3): 159 76. The key contribution on product dif ferentiation by country of origin.

Bergstrand, Jeffrey H. 1985. "The Gravity Equation in International Trade: Some Microeconomic Founda tions and Empirical Evidence." *Review of Economics and Statistics* 67 (3): 474 81. A second attempt to provide theoretical foundations to the gravity model.

. 1989. "The Generalized Gravity Equation, Monopolistic Competition, and the Factor Proportions Theory in International Trade." *Review of Economics and Statistics* 71 (1): 143 53. An interpretation of the gravity model in terms of monopolistic competition.

Brun, Jean François, Céline Carrère, Patrick Guillaumont, and Jaime de Melo. 2005. "Has Distance Died? Evi dence from a Panel Gravity Model." *World Bank Eco nomic Review* 19 (1): 99 120. A useful exploration of distance in gravity models.

Deardorff, Alan V. 1998. "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?" In *The Regionalization of the World Economy*, edited by Jeffrey A. Frankel. Chicago: University of Chicago Press, 7 22. A helpful review and assessment of the gravity model.

Feenstra, Robert C. 2004. *Advanced International Trade: Theory and Evidence*. Princeton, NJ: Princeton Uni versity Press. Chapter 5 of this book reviews the gravity model in light of trade theory and also delves into a number of key estimation issues.

Feenstra, Robert. C., James R. Markusen, and Andrew K. Rose. 2001. "Using the Gravity Equation to Differ entiate among Alternative Theories of Trade." *Canadian Journal of Economics* 34 (2): 430 47. Tests the gravity model over differentiated and homogeneous goods, fo cusing on differences in estimated parameter values.

Haveman, Jon, and David Hummels. 2004. "Alternative Hypotheses and the Volume of Trade: The Gravity

Equation and the Extent of Specialization." *Canadian Journal of Economics* 37 (1): 199 218. Explores gravity model explanations both in terms of complete speciali zation such as in monopolistic competition models and incomplete specialization with trade frictions.

Helpman, Elhanan. 1987. "Imperfect Competition and International Trade: Evidence from Fourteen Industrial Countries." *Journal of the Japanese and International Economies* 1 (1): 62 81. A claim for monopolistic competition models of intraindustry trade using gravity model evidence.

Redding, Stephen, and Anthony J. Venables. 2004. "Eco nomic Geography and International Inequality." *Jour nal of International Economics* 62 (1): 53 82. An appli cation of the gravity framework to economic geography.

Tinbergen, Jan. 1962. *Shaping the World Economy: Sugges tions for an International Economic Policy*. New York: Twentieth Century Fund. The first use of a gravity model to analyze international trade flows.

KENNETH A. REINERT

■ greenfield investment

See mergers and acquisitions

■ Group of Seven/Eight (G7/G8)

The Group of Seven/Eight (G7/G8) is an informal intergovernmental institution for the leaders of the world's most powerful countries to govern the global economy and promote the values of open democ racy, individual liberty, and social advancement. It emerged in 1973 when the finance ministers of the United States, France, Britain, and Germany, soon joined by Japan, met secretly as a Group of Five (G5) in the library of the White House to discuss a re placement for the international monetary regime of gold-linked, fixed exchange rates, which had been destroyed by unilateral American action on August 15, 1971. When the finance ministers of France and Germany became leaders, they invited their coun terparts in the G5 and Italy to Rambouillet, France, on November 15 17, 1975, to deal with the eco nomic stagflation and broader crisis of governability arising from the 1973 oil shock, the deadlocked Tokyo Round of trade negotiations, and security threats in the Middle East, Southern Europe, India, and Vietnam. They addressed economic growth, inflation and employment, international finance, trade, energy, and economic relations with the de veloping South and communist East.

At their subsequent summits, held for two or three days in late spring or summer every year, they added Canada in 1976, the European Community (later the European Union) in 1977, and the Russian Federation in 1998, to make it the G8. They have invited as partial participants the Soviet Union and then Russia since 1991; leading African and Middle Eastern democracies since 2001; China, India, Bra zil, Mexico, and South Africa since 2003; and the executive heads of multilateral organizations such as the United Nations, the International Monetary Fund, and the World Bank starting in 1996.

The G5 finance ministers continued to meet on their own and with their leaders at the summit. In 1986 Italy and Canada were added to the finance ministers to form the G7, which soon replaced the G5 and was separated from the now leaders-only summit in 1998. Usually meeting four times a year, the G5/G7 finance ministers have had some signifi cant achievements: the 1985 Plaza Accord, which produced a soft landing for the overvalued U.S. dollar, the 1987 Louvre Accord to manage the new exchange rate relationships, and action to contain the Asian-turned-global financial crisis that afflicted Thailand, Indonesia, South Korea, Russia, the Uni ted States, Brazil, Argentina, and Turkey from 1997 to 2001. Some see G7 finance performance as in herently harmful or declining during the globaliza tion of the 1990s. Others highlight its continuing importance as a source of new ideas.

Other G7/G8 ministerial bodies were created, including one for trade in 1982 and one for devel opment in 2002. In 1999, a broader Group of Twenty (G20) finance ministers and central bank governors of systemically important countries was created. At the official level, groups on energy pro liferated in the 1970s. Two on macroeconomics

arose in 1982. Many on financial crime, financial assistance to Russia, the Gulf States, development finance, Africa, terrorist finance, investment, and intellectual property have emerged since 1989.

The G7/G8 summit's economic agenda soon extended from macroeconomics into many specific microeconomic and social policy subjects. It also quickly embraced global issues such as migration, crime, drugs, and political-security subjects such as East-West relations, terrorism, proliferation of weapons of mass destruction, regional security, and the use of force. Its unique value has been in freely and flexibly addressing any and all issues, and integrating them in innovative ways. From its deliberations have flowed major new directions, such as redefining the relationship between inflation and unemployment. It has issued an expanding number of often ambitious collective commitments. Its members have generally complied at least partly with these commitments, especially in energy, trade, financial assistance to Russia, and debt relief for poor states.

The G7/G8 summit has made an effective contribution to global economic governance in several critical cases. In finance it mobilized the assistance necessary to sustain Russia's democratic revolution in the 1990s. In trade it provided a catalytic political push to launch multilateral trade negotiations in the Uruguay Round and Doha Development Agenda, and successfully conclude the overdue Tokyo and Uruguay Rounds. In energy it stopped the inflation and wealth redistributions brought by the second oil shock in 1979, although at the cost of a severe recession within the G7 and the default of several resource-rich developing states. In development, from the 1988 Toronto Summit to the 2005 Gleneagles Summit, it canceled the debt of the deserving poorest countries of the world and mobilized major new monies for development assistance. After the terrorist attack on the United States on September 11, 2001, it instituted an effective regime to control terrorist finance.

Some see G8 governance as propagating neoliberal values and perpetuating inequality between its rich Northern members and the poor developing South. Others regard it as an important source of political leadership and collective management, most clearly in its 1975 creation of a new monetary regime of managed floating exchange rates and its 1978 big package deal that integrated fiscal, monetary, trade, and energy concerns. Still others argue that it has emerged as the effective center of global governance in the post Cold War, rapidly globalizing world, where even the most powerful economies are vulnerable to shocks that require concerted action to address.

The G7/G8 is increasingly reaching out to incorporate the growing economic powers of China, India, Mexico, Brazil, and South Africa more fully in its annual summit and ministerial processes. It is reaching down to involve business and civil society in its work. Although it still lacks any formal charter or secretariat, its comprehensive reach, freedom, flexibility, direct connection with leaders, and increasing inclusiveness enhance its relevance in guiding the global economy in the 21st-century world.

See also Doha Round; globalization; Louvre Accord; Plaza Accord; Uruguay Round

FURTHER READING

Baker, Andrew. 2006. *The Group of Seven: Finance Ministries, Central Banks, and Global Financial Governance.* New York: Routledge. An analysis of the discussions within the G7 finance ministers' forum.

Bayne, Nicholas. 2005. *Staying Together: The G8 Summit Confronts the Twenty First Century.* Aldershot, UK: Ashgate. A detailed analysis of the G8 summit since 1998, offering a framework for assessing performance. It argues that G8 effectiveness is rising, due to the reforms instituted in 1998 and the G8's character as a club in which all members, not only the United States, now can and do lead.

Bergsten, Fred, and Randall C. Henning. 1996. *Global Economic Leadership and the Group of Seven.* Washington, DC: Institute for International Economics. An evaluation of the G7 finance ministers' forum, arguing that its performance has declined because its members mistakenly believe that globalization has now made markets, not their governments, the governors of the global economy.

Feldstein, Martin. 1988. "Distinguished Lecture on Economics in Government: Thinking about International Economic Coordination." *Journal of Economic Perspectives* 2 (spring): 3 13. The classic argument that G7 economic coordination is ineffective and even harmful, as G7 governments do not know the best targets or instruments to reach them any better than market players themselves.

Fratianni, Michele, Paolo Savona, and John J. Kirton, eds. 2005. *New Perspectives on Global Governance: Why America Needs the G8.* Aldershot, UK: Ashgate. A collection based on the U.S. hosted summit in 2004, arguing that a newly vulnerable America needs, uses, and adjusts to the G8 to protect itself in a post September 11 world.

Kaiser, Karl, John J. Kirton, and Joseph Daniels, eds. 2000. *Shaping a New International Financial System: Challenges of Governance in a Globalizing World.* Aldershot, UK: Ashgate. An examination of the G7/G8 response to the 1997 99 financial crisis, generally concluding that the G7/G8 was effective in the end.

Putnam, Robert D., and Nicholas Bayne. 1987. *Hanging Together: Cooperation and Conflict in the Seven Power Summits.* Rev. ed. Cambridge, MA: Harvard University Press. The classic description and systematic explanation of G7 summit performance, arguing that American leadership was a necessary condition for G7 success.

JOHN KIRTON

■ growth in open economies, neoclassical models

The neoclassical growth model is a widely used framework to study economic growth. The main focus of neoclassical growth theory is on explaining economic growth via the accumulation of factors of production, namely physical capital.

When economies are closed to foreign exchange, a central prediction of the theory is that among similar countries (those converging to the same steady-state, determined by factors such as the quality of institutions, overall level of education, government policies, etc.), the poorest should grow faster than the richest. This notion of *conditional convergence* is important

because it suggests the existence of automatic forces that pull economies closer together. Several studies, relying on either cross-region or cross-country growth regressions, provide empirical support for conditional convergence. The assumption of closed economy is, however, very difficult to justify. A first problem is that, if conditional convergence holds among closed economies, the theory implies that foreign investors must be giving up extraordinary investment opportunities. Capital should actually be flowing across countries to arbitrage such return differences. A second problem is that many economies are in fact open. This includes South Korea, Taiwan, Hong Kong, and Singapore, the so-called East-Asian miracles that achieved very rapid and sustained postwar growth. Moreover, these are well-known examples of export-oriented growth. Open economy versions of the neoclassical growth model are able to deal successfully with both challenges.

Assumptions of the Theory Consider a typical economy, potentially open to trade with the rest of the world. The analysis rests on assumptions about the production technology and the preferences of individual consumers underlying saving behavior.

The economy produces a single homogeneous good, the model's counterpart to real gross domestic product (GDP) in the data. The technology available to produce this good is described by an aggregate production function the maximum output that is technologically feasible for the economy as a whole to produce given the available factors of production. A common specification is the Cobb-Douglas production function, $Y_t = A_t K_t^{\alpha} L_t^{1-\alpha}$, where Y_t is the output (real GDP) in the current period, and K_t and L_t are the inputs, aggregate physical capital (all the plants and equipment) and aggregate labor (total number of workers), respectively. Capital may be accumulated over time; however, for simplicity, we assume L_t is constant at L. The parameter $0 < \alpha < 1$ captures the relative importance of capital in production. A_t is the level of technology, that is, the general knowledge available to society on how to produce. It captures how efficient an economy is in combining the available production inputs. Its determinants are not just narrowly technological but

also more broadly institutions and government policies. Neoclassical theory is typically not explicit about these determinants; rather it assumes A_t grows at some constant rate common to all countries. For simplicity we assume here that A_t is constant at $A > 0$.

Since we are often interested in understanding the growth of real GDP per capita, rewrite the production function as $y_t = A k_t^\alpha$, where $y_t = Y_t/L$ and $k_t = K_t/L$. This production technology has two important properties. First, since $\alpha > 0$, increasing k_t always increases production. Second, since $\alpha < 1$, increasing k_t becomes less productive when the scale of production, as proxied by k_t, is already large — a key property called decreasing returns to scale. The idea behind decreasing returns to scale is that, as k_t increases, each worker needs to operate more units of capital, and so capital becomes less productive — for example, some machines become idle sometimes, because there are not enough workers to operate all of them simultaneously.

The economy's residents have a preference for current versus future consumption. In addition, government policy such as capital income taxation influences individual saving behavior. For simplicity, we assume preferences and policy lead to a simple aggregate saving rule: national saving is a constant fraction of national income, or real gross national product (GNP). Real GNP is the residents' total income irrespective of origin — GDP plus the net income generated abroad.

The simple saving rule determines the level of saving in each period. However, the economy also needs to decide how to invest each period's saving, either domestically or abroad. We assume investors choose the option with the highest rate of return.

When today's saving is invested domestically, it will increase the capital stock tomorrow — assuming it takes one period to install new capital. Let's say capital tomorrow increases by one unit. The gross return from this investment is the extra production generated tomorrow by the additional unit of capital — what economists call the marginal product of capital, which in the present case equals $\alpha A k_{t+1}^{\alpha-1}$. Because, due to wear and tear, capital depreciates physically during production, the net return will be lower. With a rate of physical depreciation every period of $0 < \delta < 1$, only a fraction of the initial unit of capital survives production. The net return from investing domestically becomes $r_{t+1} = \alpha A k_{t+1}^{\alpha-1} - \delta$. Importantly, because of decreasing returns to scale ($\alpha - 1 < 0$), a higher k_{t+1} lowers r_{t+1}: richer countries have a lower incentive to accumulate capital because they already have more of it.

Today's saving may instead be invested abroad. We assume there is a frictionless world credit market, where any economy can borrow or lend funds freely at the going world interest rate r_{t+1}^*. Suppose every economy is a small open economy, in the sense that their individual actions have a negligible effect on the world market. For simplicity, assume also that the world interest rate is constant at r^*. Picking the investment alternative yielding the highest return means comparing r^* with r_{t+1}.

So far, with a single homogeneous good being produced in every country, openness to trade only takes place in world credit markets. That is, all trade is intertemporal, involving only borrowing and lending — so far there is no scope for intratemporal trade in different commodities.

The Benchmark Case of a Closed Economy A closed economy has no access to world credit markets, so national saving is all invested domestically. Equivalently, domestic investment in physical capital must be financed through national saving.

Suppose the economy is relatively poor; in particular the current levels of capital and production are lower than their steady-state values. Because of decreasing returns to scale, r_{t+1} is high. With a high return, the saving rule generates enough investment to compensate for depreciated capital and still increase the capital stock tomorrow. Hence, the economy will be growing. However, the growth rate will diminish over time, since decreasing returns to scale reduce returns as capital grows. At some point, we reach a situation where the economy invests just the amount of resources needed to cover depreciation. At that point — the steady-state — capital and output per capita are constant over time, and the economy stops

growing. In the more general case in which the rate of technological progress is positive (A_t grows over time), then GDP per capita would grow at this same rate in the steady-state.

Convergence to the steady-state from the initial level of GDP is only gradual. Convergence is not faster because domestic saving is insufficient to generate all the investment necessary to close the gap between the current and the steady-state levels of capital per capita.

Neoclassical growth theory views cross-country differences in growth rates as stemming from countries gradually converging toward their steady-states, starting from different initial levels. Poor countries, induced by higher returns, grow faster. As Lucas (1990) pointed out, this poses a quantitative challenge to the closed economy assumption. Rewrite the (gross) domestic return as $r_{t+1} + \delta = \alpha A^{1/\alpha} y_{t+1}^{(\alpha-1)/\alpha}$, using the definition of y_t. Consider now the actual experiences of the United States and Kenya through the lens of neoclassical theory. If both countries shared the same technology and differed only in their current position as they transit to a common steady-state, then

$$\frac{r^{KEN} + \delta}{r^{US} + \delta} = \left(\frac{y^{KEN}}{y^{US}}\right)^{(\alpha-1)/\alpha}$$

In 2003, the United States was about 30 times richer than Kenya. With $\alpha = 1/3$, as suggested by the data, then investing in Kenya would have earned a gross return $30^2 = 900$ times higher than in the United States! Even if we attempted to factor in the technological differences that do exist between Kenya and the United States, the return gap would still come out too high. The incentive for capital to flow from the United States into Kenya would be enormous, making it impossible to sustain the closed economy assumption.

Openness to Capital Flows The small open economy has free access to the world capital market, at the constant interest rate r^*. This interest rate reflects the overall return on capital in the rest of the world. Consider a scenario in which the economy is initially in autarky, and suddenly gains free access to the world credit market. Assume the domestic economy is poorer than the rest of the world. Due to decreasing returns to scale, domestic returns under autarky are higher than the world interest rate, $r_{t+1} > r^*$.

Faced with this return gap, foreign investors have an incentive to invest in the domestic economy. There will be a massive instant inflow of funds, which fuels domestic investment. As the stock of capital increases, r_{t+1} declines. The inflow of funds only ceases when $r_{t+1} = r^*$. The domestic economy reaches the steady-state instantly, and GDP automatically attains the level of rich countries. The economy forever becomes a net debtor in world markets, with external debt being serviced via the permanent increase in real GDP.

Differently from the closed economy case, convergence to the steady-state is extremely rapid. Convergence can be fast now, since investing in domestic production may be in part financed with foreign funds, and it does not have to rely only on domestic saving.

Slow Convergence with International Capital Flows Infinite convergence speeds are clearly counterfactual. Realistic convergence speeds obtain if the basic model is extended as in Cohen and Sachs (1986). First, we may relax the assumption that the domestic economy has unrestricted access to foreign funds. In order to finance domestic investment, assume foreign investors require residents to use a certain fraction of their own resources along with foreign funds. This might occur if foreign investors fear residents might default on their debt in the future. Cohen and Sachs (1986) have introduced such credit constraint in the basic model. Barro, Mankiw, and Sala-i-Martin (1995) have introduced a similar type of constraint in an augmented version of the basic model, where not only physical capital but also human capital is an input to production. They assumed that while foreign investors impose no constraints when financing the accumulation of physical capital, they are unwilling to finance the accumulation of human capital — that is, investment in education. The idea is that while physical capital may be offered as collateral on foreign borrowing, human capital cannot be — on default, physical capital is more easily

repossessed by foreign investors than human capital. Since countries need to accumulate both physical and human capital in order to grow, a constraint on human capital accumulation is also a constraint on physical capital accumulation. Both types of constraint on borrowing generate realistic convergence speeds for credit-constrained economies.

Second, we may assume the existence of adjustment costs to investment in domestic capital. Suppose installing new capital entails costs that are higher than just the price of capital, and that these costs become much higher for very large investment rates. Then, the gap between r_{t+1} and r^* no longer determines by itself the incentives to invest: even if $r_{t+1} > r^*$, the effective return on domestic capital may well fall below r^* for large enough levels of investment. Economies will have to adjust slowly toward the steady-state, at realistic speeds.

Openness in Goods Markets To allow for trade in commodities, we need to extend the basic one-good model to multiple goods. Such a departure has been considered by Ventura (1997). Instead of the technology described previously, suppose real GDP is produced using two intermediate inputs call them traditional and modern. Both are produced using capital and labor, but the modern commodity is more capital intensive. A closed economy behaves exactly as the basic one-good neoclassical growth model described above the more general production technology leaves the convergence implications unaltered.

Suppose instead that both commodities are freely traded, while real GDP is nontraded, and capital and labor are immobile internationally. Due to free trade, commodity prices in the small open economy equal the corresponding world prices. With capital and labor perfectly mobile across the production of the two commodities, factor prices are also constant across countries sharing the same technology an instance of the factor-price-equalization theorem. That is, even in the absence of international factor movements, wages and the returns to capital are indirectly pinned down by trade in goods, namely we get $r_t = r^*$.

If saving rates are constant, like before, then not only rich and poor countries wish to invest at the same rate, now they also enjoy the same return r^*. Since trade removes the poor countries' edge in terms of high returns, they stay poor relative to rich countries. There is neither (conditional) convergence nor divergence in world incomes: two economies with identical production technologies will forever stay different if they start out the growth process different. This stands in sharp contrast with both the closed economy implications and those of the basic one-good model.

Whether GDP actually converges or diverges across countries depends on the behavior of saving rates over time. If saving rates increase with the level of GDP, then incomes diverge over time. For convergence still to take place with trade, saving rates must decrease as a country develops.

The East-Asian Miracles Trade in goods also helps us understand the extraordinary growth experience of the East-Asian miracles. The basic one-good model, as well as the multiple-good model without trade, has trouble explaining very high growth rates, sustained over long periods of time: it requires not only high but also increasing saving rates, capable of defying decreasing returns to scale. A much more plausible explanation is that, because of trade-induced return equalization, economies may sustain very high growth rates by avoiding decreasing returns altogether. The performance of the East-Asian miracles may then be understood by the success of these economies at raising saving rates in the early 1960s, and their heavy reliance on exporting industries. Instead of meeting decreasing returns, the fast rates of capital accumulation were channeled to the modern capital-intensive sector. This sector expanded at the expense of the traditional one, its increased production heavily exported in exchange for traditional good imports. In a nutshell, the East-Asian miracles were successful by having the rest of the world absorb their high capital accumulation rates through manufacturing exports.

See also economic development; growth in open economies, Schumpeterian models; international income convergence; trade and economic development, international

FURTHER READING

Barro, Robert J., N. Gregory Mankiw, and Xavier Sala i Martin. 1995. "Capital Mobility in Neoclassical Models of Growth." *American Economic Review* 85 (1): 103 15. Argues that incorporating human capital reduces the rate of convergence in the open economy neoclassical model to a realistic level.

Barro, Robert J., and Xavier Sala i Martin. 2004. *Economic Growth*. 2d ed. New York: McGraw Hill. A compre hensive and authoritative, although technical, treatment of growth theory and evidence, starting from neoclassical models end extending beyond.

Cass, David. 1965. "Optimum Growth in an Aggregative Model of Capital Accumulation." *Review of Economic Studies* 32 (3): 233 40. A classic reference and one of the building blocks of the neoclassical growth model.

Cohen, Daniel, and Jeffrey D. Sachs. 1986. "Growth and External Debt under Risk of Debt Repudiation." *Eu ropean Economic Review* 30 (3): 529 60. Argues that fear of default will deter large international capital flows, slowing the rate of income convergence.

Jones, Charles I. 2002. *Introduction to Economic Growth*. 2d ed. New York: W. W. Norton. Includes a very acces sible exposition of growth facts and neoclassical growth theory.

Koopmans, Tjalling C. 1965. "On the Concept of Optimal Economic Growth." In *The Econometric Approach to Development Planning*. Amsterdam: North Holland, 1965, 225 300. A classic reference and one of the building blocks of the neoclassical growth model.

Lucas, Jr., Robert E. 1990. "Why Doesn't Capital Flow From Rich to Poor Countries?" *American Economic Review, Papers and Proceedings* 80 (2): 92 96. A short and accessible discussion of the international return equalization challenge facing neoclassical theory.

Ramsey, Frank. 1928. "A Mathematical Theory of Saving." *Economic Journal* 38 (152): 543 59. A classic reference and one of the building blocks of the neoclassical growth model.

Solow, Robert M. 1956. "A Contribution to the Theory of Growth." *Quarterly Journal of Economics* 70 (1): 65 94. A classic reference and one of the building blocks of the neoclassical growth model.

Ventura, Jaume. 1997. "Growth and Interdependence." *Quarterly Journal of Economics*. 112 (1): 57 84. Argues

that diminishing returns to capital is much less im portant to open economies, which helps explain the extraordinary growth of some east Asian economies.

RUI CASTRO

■ growth in open economies, Schumpeterian models

Schumpeterian growth is a particular type of eco nomic growth generated by the endogenous intro duction of product and/or process innovations. The term *endogenous* refers to innovations that result from research and development (R&D) investments un dertaken by forward-looking, profit-seeking firms. The term *Schumpeterian growth* is named for Joseph Schumpeter, who described the evolution of capi talism through a process of "creative destruction." This process serves as the fundamental building block in models of Schumpeterian growth and cap tures the social costs and benefits that result from the endogenous destruction of old technologies by new ones: "Economic progress, in capitalist society, means turmoil. And . . . in this turmoil competition works in a manner completely different from the way it would work in a stationary process, however per fectly competitive. Possibilities for gains to be reaped by producing old things more cheaply are constantly materializing and calling for new investments. These new products and new methods compete with the old methods not on equal terms but at a decisive advantage that may mean death to the latter" (Schumpeter 1942, 42). Schumpeterian growth models are closer in spirit to Schumpeter's ideas than are other theoretical approaches to economic growth such as those that emphasize learning-by-doing, human capital accumulation, or physical capital ac cumulation as sources of economic growth.

The development of Schumpeterian growth the ory started in the early 1990s, motivated by the di vergence of national growth rates, Japan's challenge to the United States' technological leadership, and the inability of the neoclassical growth theory to ac count for the long-run causes of technological progress. The New Growth Theory has primarily

focused on causes and effects of long-run technological progress, which is the sole determinant of long-run growth of income per capita. One of the most important insights of the New Growth Theory is the relationship between the economics of ideas and economic growth. Two inherent features of ideas are that they are *nonrival*: an operations manual for organizing a retail store can be used by many stores without any significant replication costs other than photocopying the manual, whereas consumption of an apple by a consumer precludes its consumption by another consumer. Ideas are also partially *excludable* in the sense that the owner of an idea can charge a fee for its use because he/she can restrict access to other users. For instance, Coca-Cola can restrict access to its formula, at least temporarily, by locking the recipe in a safe or by acquiring a patent.

Nonrivalry and partial excludability imply that ideas, which fuel the generation of technology, need to be produced only once, and involve a fixed cost and negligible marginal production cost. Consequently, the economics of "ideas" is necessarily related to the presence of increasing returns to scale and imperfect competition. In a general-equilibrium framework, temporary monopoly profits are essential to finance the R&D cost that occurs before manufacturing of new products (based on new ideas) takes place. Therefore, monopoly profits are not necessarily bad for society but provide incentives for firms to engage in risky R&D investments to discover new products or better processes. The New Growth Theory shifts its focus from competition in product-markets to intertemporal competition in the market for innovations. The latter determines the rate of product development, which generates long-run Schumpeterian growth and limits the duration of monopoly power in product markets.

One variant of Schumpeterian (R&D-based) growth theory focuses on economic growth generated by the introduction of better-quality products. These dynamic general-equilibrium models capture the process of creative destruction as follows: first, sequential and stochastic R&D races are typically used to formalize the entrepreneurial risk, uncertainty, and fixed cost that are inherent in the innovation process; second, product obsolescence, which is based on the endogenous replacement of old, lower-quality goods by new, higher-quality ones, formalizes the Schumpeterian notion of creative destruction; and third, the assumption of patents granted to inventors of new goods formalizes the role of Schumpeterian temporary monopoly power that fuels technological change.

The presence of increasing returns and temporary monopoly power implies a divergence between the market and social optimum and leaves plenty of room for government intervention. In general, the welfare ranking between the social and market rates of Schumpeterian growth is ambiguous thanks to the presence of a few forces that create a divergence between social and private incentives to innovate. One force has been christened the *monopoly-distortion* effect. This force creates an incentive for overinvestment in R&D relative to the social optimum: firms are motivated by the instantaneous monopoly profit margin when engaged in R&D, which happens to exceed the increase in social value (measured by consumer surplus) generated by a typical innovation. Another countervailing force has been christened the *intertemporal-spillover* effect. This force creates a tendency for underinvestment in R&D relative to the social optimum: the social planner takes into account the fact that consumers benefit forever from an innovation. In contrast, being aware that their lives are finite (due to the process of creative destruction), private firms engaged in R&D discount the returns to innovation more than the social planner and invest less. This welfare ambiguity is present in both closed and open-economy Schumpeterian growth models.

First-generation models of Schumpeterian growth exhibit a counterfactual "scale-effects" property according to which more resources devoted to R&D are associated with a higher growth rate of total factor productivity (TFP), that is, output growth not accounted for by the growth in inputs, and the presence of a positive population growth rate must generate an unbounded (infinite) long-run growth rate of per capita income. In other words, according to these models, if R&D employment doubles, then the rate

of product creation and the economy's growth rate must also double. Since 1950, however, R&D employment in the United States and other advanced countries has more than doubled without generating any significant acceleration in TFP growth.

Since the mid-1990s, growth theorists have developed a second generation of scale-free, Schumpeterian growth models that fall into two distinct categories depending on the way of removing the scale effects property. Semi-endogenous Schumpeterian growth models incorporate diminishing returns to the stock of knowledge that affects the productivity of R&D resources by assuming that, as technology becomes more complex, sustained growth in R&D resources is needed to maintain a given rate of TFP growth. In other words, easier inventions are discovered first, and as the set of technological opportunities diminishes over time, it becomes more difficult to discover new products. Semi-endogenous growth theory predicts that the long-run growth rate of TFP depends only on the rate of population growth, and therefore it is not affected by policy-related parameters. In other words, semi-endogenous growth theory generates exogenous scale-free long-run Schumpeterian growth.

The second category of R&D-based growth theory consists of fully endogenous and scale-free Schumpeterian growth models. This approach to the removal of scale effects builds on the insight that aggregate R&D effort becomes less effective either because it is spread among more product lines (as new varieties are discovered), or because incumbents who face a risk of going out of business raise barriers to frustrate the R&D effort of challengers. Fully endogenous Schumpeterian growth models maintain the assumption of constant returns to the stock of knowledge of earlier endogenous growth models and generate endogenous long-run growth. In other words, with the exception that the size of an economy measured by its population level does not affect long-run growth, these models share the same properties as earlier endogenous growth models. Although the evidence against the empirical validity of earlier endogenous growth models is convincing, the ongoing debate on whether semi-endogenous or

fully endogenous Schumpeterian growth models are more empirically relevant is inconclusive. The terms *earlier*, *semi*, and *fully* endogenous growth models will be used to refer to first-generation, scale-free exogenous, and scale-free endogenous Schumpeterian growth models, respectively, to conserve space and to clarify the exposition.

Economic Openness and Growth Open-economy models of Schumpeterian growth constitute the backbone of dynamic trade theory and complement traditional trade theory by focusing on the analysis of the economic forces that determine the generation and international transfer of technology. They provide valuable theoretical insights into the patterns of growth and trade, global income distribution and poverty, and the effects of economic policies on the performance of the global economy.

Earlier Schumpeterian growth models analyzed the nexus between trade patterns and long-run growth using a variety of approaches. They generated product-cycle trade, which is based on the observation that many products are first discovered in advanced countries and then their production shifts to less advanced countries as their technology is imitated, within a context of a high-wage innovating North and a low-wage noninnovative South, or within the context of two Northern economies facing equal factor prices but differing in their factor endowments. They also identified the economic determinants of sustained comparative advantage in high-technology industries: countries with higher comparative labor productivity in R&D activities, or countries that are abundant in factors used intensively in R&D, such as human capital, will export high-tech products to the rest of the world.

Schumpeterian models have also analyzed the determinants of growth in open economies. Earlier endogenous growth models identified three broad channels through which economic openness affects long-run growth. First, trade, by increasing the size of the market, raises the profitability of R&D investment and increases the long-run rate of innovation and growth in all trading countries. Second, economic openness, by facilitating the international exchange of information, increases the scope of

knowledge spillovers, raises the productivity of researchers, and accelerates the rates of innovation and growth in the global economy. Third, trade openness, by reallocating economic resources across sectors and between R&D investment and manufacturing activities, affects the long-run rates of innovation and growth. This trade-induced "specialization" process, which is the major source of welfare gains in static models, has an ambiguous effect on long-run growth.

It should be noted, however, that the market-size-based effect of trade on growth can be traced to the scale-effects property of earlier endogenous growth models. Fully endogenous growth models predict more moderate effects of economic openness on growth. These effects operate through changes in relative prices and per capita reallocation of resources between R&D and manufacturing activities. For example, a reduction in international trade costs can accelerate the long-run rate of growth by increasing the flow of temporary monopoly profits and by shifting per capita resources from manufacturing production (that is, consumption) to R&D activities (that is, investment).

Globalization and Income Distribution Schumpeterian models of economic growth have provided valuable theoretical insights associated with the evolution of wage and income inequality within and across countries. The 1970s and 1980s witnessed an alarming rise in the demand for skilled labor and a decline in the demand for unskilled labor. These labor-market developments raised the relative wage of skilled workers in the United States and several developing countries and increased the unemployment rate of less-skilled workers in several advanced countries. Economic openness and skilled-biased technological change have been proposed as two competing explanations for the rise in the relative demand for skilled labor. In traditional trade theory, relative commodity prices represent the only channel that transmits the effects of trade openness on wages. The trade-based explanation for the rise in U.S. wage inequality has been rejected by evidence on U.S. relative prices, which have remained more or less constant during the period of rising wage inequality.

Unlike traditional trade theory, Schumpeterian growth theory views technological change as an endogenous process that could be affected by changes in the reward to innovation. The reward to innovation is proportional to the expected discounted profits associated with the discovery of new products. An increase in trade openness caused by a reduction in international trade costs can affect the profitability of R&D without necessarily causing any change in relative commodity prices. Under the reasonable assumption that R&D is a skilled-labor-intensive activity, Schumpeterian growth models predict that globalization increases the relative demand for skilled labor and accelerates the rate of technological progress. An increase in the relative demand for skilled labor can generate either a rise in the relative wage of skilled workers or a rise in the unemployment rate of less-skilled workers depending on the degree of labor-market flexibility. This novel mechanism has been formally established in the context of semi-endogenous and fully endogenous Schumpeterian models of economic growth.

The controversial issue of the dynamic effects of various components of globalization on the income gap between advanced (Northern) and poor (Southern) countries has drawn the attention of Schumpeterian growth theorists. An important strand of Schumpeterian models of North-South trade and growth has analyzed the income distributional effects of globalization-enhancing policies. The main insight of this literature is that any increase in the rate of imitation and/or decline in the rate of innovation reduces the North-South wage inequality measured by the Northern relative wage. In other words, faster international transfer of technology from North to South benefits Southern workers.

Government Interventions Schumpeterian models of economic growth have analyzed the long-run growth and welfare effects of a variety of government interventions. Policy instruments such as taxes and subsidies on R&D, production, and trade change relative product and factor prices and generate shifts in economic resources between consumption and R&D activities. In earlier and fully endogenous growth models, a policy-induced shift in per capita

resources toward R&D activities accelerates permanently the rates of innovation and growth. In semi-endogenous growth models this resource shift generates a temporary (as opposed to a permanent) increase in the rate of innovation. In the presence of asymmetric industries and countries, dynamic general equilibrium forces can reverse the well-intended effects of several policy interventions. For instance, R&D subsidies may cause a country to export fewer R&D-intensive goods and import more of them; and industrial policies that subsidize manufacturing activities in high-technology sectors may have detrimental effects on global long-run innovation and growth because they could raise the costs of R&D by shifting resources away from R&D activities.

Several North-South Schumpeterian growth models have analyzed the dynamic effects of stronger protection of intellectual property rights (IPR). Stronger IPR protection, modeled either as an increase in patent length or as a reduction in the rate of Southern imitation, reduces the rate of international technology transfer from innovating North to imitating South, raises the North-South wage gap, and has an ambiguous effect on the rate of innovation and global Schumpeterian growth. Even if stronger IPR protection increases the rate of Northern innovation in the short run, it could reduce the welfare of Southern consumers and could raise the welfare of Northern consumers by shifting production from low-price South to high-price North. Consequently, North-South models of Schumpeterian growth are well fitted to provide useful policy recommendations regarding the intense debate surrounding the Agreement on Trade-Related Intellectual Property Rights (TRIPS), which was ratified by the members of the World Trade Organization in 1995 and calls for stronger Southern IPR protection.

See also Agreement on Trade-Related Intellectual Property Rights (TRIPS); North-South trade; technological progress in open economies

FURTHER READING

Aghion, Philippe, and Peter Howitt. 1998. *Endogenous Growth Theory*. Cambridge, MA: MIT Press. Provides a comprehensive analysis of endogenous growth theory.

Dinopoulos, Elias, and Peter Thompson. 1999. "Scale Effects in Schumpeterian Models of Economic Growth." *Journal of Evolutionary Economics* 9 (2): 157 85. Reviews available empirical evidence on the relationship between scale and growth, and attempts to develop Schumpeterian growth without scale effects.

Grossman, Gene M., and Elhanan Helpman. 1991. *Innovation and Growth in the Global Economy*. Cambridge, MA: MIT Press. Presents a coherent theoretical framework that integrates the variety based and quality based approaches of earlier Schumpeterian models of growth in open economies.

Schumpeter, Joseph. A. 1942. *Capitalism, Socialism, and Democracy*. New York: Harper and Row. This classic book provides a comprehensive treatment of the process of creative destruction.

Segerstrom, Paul S., T.C.A. Anant, and Elias Dinopoulos. 1990. "A Schumpeterian Model of the Product Life Cycle." *American Economic Review* 80 (5): 1077 91. This is the first model of Schumpeterian growth in open economies.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5, part 2): S71 S102. This seminal article developed the most influential and most widely cited closed economy model of Schumpeterian growth.

ELIAS DINOPOULOS

■ **Grubel-Lloyd index**
See intraindustry trade

■ **Gulf Cooperation Council**

The Gulf Cooperation Council (GCC) was created in February 1981 by six Arab Gulf states: Bahrain, Kuwait, Qatar, Oman, Saudi Arabia, and the United Arab Emirates (UAE). The main motive for the creation of the GCC was to face the threat posed to the region's security by the Iran-Iraq war at that time. The aims of the GCC formation are to develop cooperation and integration among the member states on foreign and defense policies and to promote

common interest in economic, social, and cultural affairs. A customs union that allows no tariffs or other barriers to trade among members (as in free trade areas) and harmonizes trade policies (such as setting of common tariff rates) toward the rest of the world was created between GCC members in January 2004.

Members of the GCC share common features such as religion, language, historical background, and social life. The GCC economies depend on similar natural resources and have a comparable structural base. The GCC members also share the same sociopolitical system in which the states, ruled by kings and princes, own most of the natural resources and the large public sector dominates all aspects of the economy.

The GCC population as of 2005 was approximately 36 million. Saudi Arabia has the largest population base compared to the other five members. Approximately 38 percent of the total GCC population is composed of expatriates with no permanent residence. The proportion of nationals out of the total population varies from one country to another. This ratio is very small in Kuwait, Qatar, and the UAE (35, 27, and 22 percent, respectively), compared to Bahrain, Oman, and Saudi Arabia (62, 78, and 72 percent). Even though the GCC fertility rate is among the highest in the world, their economic dependence on foreign labor is not decreasing.

GCC social indicators depict a very positive human development profile. Life expectancy is well above 70 years, infant mortality is far below the world average, and the population per physician is very low. However, fertility rates in most GCC countries are among the highest in the world, reflecting generous family allowances provided by governments in order to increase the population base. Finally, education spending has contributed to declining illiteracy rates.

Expatriates constitute the largest percentage of the labor force in GCC countries. Table 1 shows that approximately one-third of the total labor force in GCC countries are citizens while two-thirds are expatriates. The percentage of citizen labor force in the states of Kuwait and the UAE is less than 20 percent. Only in Saudi Arabia is the percentage of citizens in the labor force slightly greater than 50 percent of the total labor force. The greatest majority of GCC employees are working in the public sector, and the average wages and salaries of GCC citizens are almost three times higher than those of expatriates.

GCC economies depend entirely on the production of oil, and all economic sectors revolve around revenues generated by oil exports. The exploration for oil began in the Gulf region in 1945. The Gulf oil has been important in the global energy market for many reasons. First, the oil reserves of the GCC are huge in comparison to the world's total reserves. The GCC oil reserves constitute around 50 percent of the world total. Second, the GCC plays a significant role in the supply of oil to the world market. The share of the GCC in total world production is approximately 25 percent. Third, geological factors such as the location of the onshore oil fields close to the deep Persian Gulf, the flow of the oil toward the sea, and the ease of drilling have helped the GCC's oil to be produced relatively more cheaply than the rest of

Table 1
The GCC labor force in 2005 (thousands)

	Bahrain (%)	Kuwait (%)	Oman (%)	Qatar (%)	SA (%)	UAE (%)	GCC (%)
Citizens	145 (41.7)	327 (19.5)	205 (31.1)	41 (44.1)	3,469 (51.3)	450 (15.7)	4,637 (37.4)
Expatriates	203 (58.3)	1,349 (80.5)	454 (68.9)	52 (55.9)	3,289 (48.7)	2,411 (84.3)	7,758 (62.6)
Total	348 (100)	1,676 (100)	659 (100)	93 (100)	6,758 (100)	2,861 (100)	12,395 (100)

Sources:
The Cooperation Council for the Arab States of the Gulf: Statistical Bulletin, 2006.
Ministry of Planning in various GCC countries.

world's oil. Finally, the central geographical location of the Persian Gulf between the developed economies in the West and growing economies of East Asia has reduced transport costs and increased the significance of the GCC oil market.

Table 2 reveals that, at the end of 2005, oil reserves in GCC countries were approximately 470 billion barrels, oil products exceeded 15 million barrels per day, and oil revenue was approximately $187 billion. Saudi Arabia is the largest oil producer in the GCC countries, followed by Kuwait and the UAE. However, Qatar and Oman have much smaller oil reserves, and much less oil production and oil revenue, than these three members of the GCC. Bahrain seems to be losing its dependence on oil.

The petroleum and mining contribution to GDP in GCC countries constitutes 20 to 60 percent of total gross domestic product (GDP). This contribution is higher when oil prices rise. The contribution was the highest in Qatar (61 percent) and the lowest in Bahrain (22 percent) in 2005. The second most important economic activity is the services sector (wholesale and retail trade, transportation and communication, finance and insurance, real estate). The services sector contributes 20 to 30 percent of GDP. The contribution of the service sector in 2005 was 51 percent in Bahrain (which is no longer a major oil producer) and 15 percent in Qatar (which depends heavily on its production of oil and natural gas). The third most important sector is the government, which contributes 10 to 20 percent of GDP. The manufacturing sector plays a moderate

role in all members of the GCC. This sector contributes only 6 to 10 percent of GDP. The only exception is its contribution in the United Arab Emirates in 2005, which amounted to 14 percent. Contribution of the agricultural sector to GDP is negligible in most GCC countries and was only 4 percent in Saudi Arabia in 2005. The GCC members heavily subsidized the state-owned electricity, gas, and water sector.

The level of per capita income differs significantly among members of the GCC. Per capita income in the state of Qatar was the highest in 2005 ($39,101), while per capita income in Oman was the lowest ($9,939). Per capita income in all GCC members was approximately $14,346 in 2005.

The proportion of total expenditure on exports and imports constituted a significant proportion of GDP. Re-exporting is a major part of the Bahrain and UAE total exports (around 30 percent). The proportion of total expenditure on private consumption out of GDP is much lower than in countries with similar per capita income (around 35 percent) and is especially low in the state of Qatar (15 percent of GDP in 2005). Expenditure on public consumption was approximately 20 percent of GDP in the states of Kuwait, Oman, and Saudi Arabia in 2005.

The percentage of GDP devoted to gross fixed capital formation in the GCC countries varies from less than 15 percent in the case of Kuwait to 25 percent in the case of the UAE. The government carries out a large percentage of investment in these

Table 2
GCC oil production, revenue, and reserves in 2005

	Bahrain	Kuwait	Oman	Qatar	Saudi Arabia	UAE	GCC
Oil reserves (billion barrels/year end)	0.12	96.8	5.8	4.5	264.5	99	470.72
Oil production (thousand of barrels/daily)	36	2,318	816	792	8,958	2,518	15,436
Oil revenue ($M)	2,700	27,340	8,312	8,124	110,654	30,124	187,254

Sources:
  The Cooperation Council for the Arab States of the Gulf: Statistical Bulletin, 2006.
  Ministry of Planning in various GCC countries.

Table 3
Total trade and intratrade of members of the GCC Customs Union (1981 2005)

	1981			1989			1997			2005		
	Total trade (US$M)	Intra trade (US$M)	% of intra trade to total trade	Total trade (US$M)	Intra trade (US$M)	% of intra trade to total trade	Total trade (US$M)	Intra trade (US$M)	% of intra trade to total trade	Total trade (US$M)	Intra trade (US$M)	% of intra trade to total trade
Bahrain	8,471	3,750	44.3	5,849	1,427	24.4	10,468	1,415	13.5	18,283	2,564	14.0
Kuwait	23,161	878	3.8	17,425	677	3.9	23,011	1,055	4.6	34,516	1,654	4.8
Oman	6,695	560	8.4	6,620	3,352	50.6	12,509	2,325	18.6	19,165	3,437	17.9
Qatar	6,907	180	2.6	3,828	330	8.6	8,376	572	6.8	21,019	1,507	7.2
SA	148,472	3,371	2.3	49,537	2,721	5.5	89,135	5,446	6.1	155,412	6,993	4.5
UAE	30,604	1,494	4.9	26,034	1,354	5.2	54,231	2,988	5.5	114,613	6,078	5.3
All GCC members	217,403	10,233	4.7	109,293	9,861	9.0	197,910	13,801	7.0	363,008	22,233	6.1

countries. Public investment has been concentrated mainly in infrastructure and the oil sector, including petrochemical industries. Private fixed capital formation, on the other hand, was directed mainly to the construction sector and the (modest) manufacturing sector. Public investment is, to a great extent, autonomous of changes in demand and interest rates and is financed from government oil revenue. Private investment depends heavily on growth in private consumption, which is greatly affected by growth in government expenditure in the GCC countries. This suggests that total (public and private) investment in the GCC countries depends on lagged oil revenues.

The GCC countries depend heavily on the outside world for the supply of most of their needs (Metwally and Tamaschke 1980; Metwally and Daghistani 1987). This is because of the relatively weak productive capacity of these economies due to lack of resources, particularly labor, materials, and water. The downturn in oil prices after the end of 1982 resulted in a sharp reduction in GCC spending on imports (Metwally 1993). However, the proportion of GDP spent on imports of goods and services in 2005 (following the rise in oil prices) was approximately 40 percent. Moreover, available statistics suggest that the income elasticity of imports exceeds one in many GCC members (Metwally 2004a, b).

Many studies have developed econometric techniques and models to identify functional relationships in the GCC economies. Metwally (1987, 1993) attempted to examine the determinants of the external surplus or gap between exports and imports of the oil-producing members of the GCC. In spite of the sharp rise in oil prices during the 1970s, the GCC countries could not improve their external surplus per exported barrel. This indicated that the rise in oil prices was greatly matched by a larger increment in imports and by the fall in the volume of exports. Metwally tested the hypothesis that the external surplus balance varies inversely with GDP. In contrast to economic theory, the results showed that the external surplus is positively correlated with GDP in the case of GCC countries. The result was explained by the fact that total GDP in the GCC countries is dominated by oil revenues, which are owned by government and are not directly available for domestic expenditure. Thus an increase in oil exports would increase total GDP and add to the overall surplus. When non-oil income was used instead of total GDP, a significant negative correlation

was obtained between non-oil income and the external surplus. Metwally also tested the interaction between the economies of the GCC and the rest of the world. The oil exports of the GCC responded favorably to the increase in the share of the Organization of the Petroleum Exporting Countries (OPEC) in world oil supply and the increase in world oil consumption. Finally, the marginal propensity to import out of non-oil income was extremely high in all GCC countries. This resulted in an "import trap," i.e., a tendency to increase imports even when the value of exports is declining (Metwally 1993; Metwally and Rammadhan 2000; Metwally 2003).

Table 3 gives data on total trade and intra-trade of members of the GCC customs union during the period 1981 2005. It is clear that the rise in oil prices resulted in higher rates of growth of intra-imports and intra-exports of most members of the GCC (Metwally and Alsowaidi 2005). Also, most commodities produced in GCC countries are not substitutes for commodities imported from outside countries. A rise in oil prices results in an increase in intra-exports due to improvement of domestic productive capacity through use of advanced technology. The analysis of Kaul, Metwally, and Perera (2007) suggests that intra-trade of most GCC members is strongly influenced by oil prices. The results also suggest that intra-exports of Qatar, Saudi Arabia, and the UAE are positively related to their intra-imports. These results also indicate that there are very significant feedback effects in intra-trade of Qatar, Saudi Arabia, and the UAE. An increase in these members' intra-imports results in an increase in their importers' incomes. However, there are no significant feedback effects in intra-trade of Bahrain, Kuwait, and Oman. These members might benefit from giving more attention to dynamic benefits of their customs union.

See also customs unions; migration, international; Organization of the Petroleum Exporting Countries (OPEC); petroleum; primary products trade; regionalism

FURTHER READING

Kaul, Nandini, M. M. Metwally, and Nelson Perera. 2007. "Feed back Effects of Intra trade between GCC Coun tries." *International Journal of Applied Business and Eco nomic Research* 5 (1): 75 89. Reveals the existence of feedback effects between trade of some members and trade of others, but not between trades of all members of the GCC.

Metwally, M. M. 1987. "Determinants of the External Surplus of the Member States of the Gulf Co operation Council." *Applied Economics* 19 (3): 305 16. Shows a significant negative correlation between the external surplus and non oil income in each GCC member country.

———. 1993. "The Effect of the Decline in Oil Revenue on the Import Patterns of the Members of the Gulf Co operation Council." *International Journal of Energy Re search* 17 (5) (July): 413 22. Proves that the decline in oil revenue has resulted in a sharp rise in income elasticity of imports for all import groups in all GCC members.

———. 2003. "Impact of Price Elasticity of Exports on Terms of Trade: The Case of the GCC Countries." *Asia Pacific Journal of Energy* 13 (1) (June): 17 24. Proves that the faster the improvement in terms of trade of GCC countries, the lower the price elasticity of demand for oil.

———. 2004a. "Impact of Fluctuations in Oil Revenue on Investment in the GCC Countries." *Economia Inter nazionale* 57 (2): 173 89. Shows that the demand for investment in GCC countries was subject to structural shifts due to fluctuations in oil prices during the last quarter of the 20th century.

———. 2004b. "Determinants of Aggregate Imports of GCC Countries: Co Integration Analysis." *Applied Econometrics and International Development* 4 (3): 59 76. Reveals that the downturn in oil prices resulted in a drastic reduction in the growth rates of imports of all the oil producers of the GCC.

Metwally, M. M., and Saif S. Alsowaidi. 2005. "Towards Unifying Monetary Policies in GCC Countries." *Global Journal of Finance and Economics* 2 (2): 149 62. Suggests that an application of a unified monetary policy for the integrated GCC countries could present difficulties.

Metwally, M. M., and A. I. Daghistani. 1987. "The Inter action between the Economies of the Member States of the Gulf Co operation Council and the Industrialized Countries." *Indian Economic Journal* 35 (3): 51 59. Reveals that despite the large volume of financial reserves

held by members of the GCC, their imports seem to depend on their export proceeds.

Metwally, M. M., and M. Rammadhan. 2000. "Impact of Fuctuations in Oil Prices on the Resource Balance of the GCC Countries." *Middle East Business and Economic Review* 12 (2): 1 9. Suggests that the resource balance of each GCC member is negatively correlated with non oil income and positively correlated with growth in the world economy.

Metwally, M. M., and R. Tamaschke. 1980. "Oil Exports and Economic Growth in the Middle East and North Africa." *Kyklos* 33 (3): 499 522. Suggests that members of the GCC need to take more measures to exploit the investment opportunities generated by the growth in oil exports.

———. 2001. "Trade Relationship between the Gulf Cooperation Council and the European Union." *Eur opean Business Review* 13 (5): 292 96. Shows significant feedback effects in GCC countries that trade with the EU.

MOKHTAR M. METWALLY

■ health and globalization

Globalization interacts with health in several areas. Globalization directly affects economic outcomes, such as the level of income per capita, that may affect health outcomes, while health may affect the level or pattern of economic activities across countries. The dissemination of health-related know-how and internationally coordinated public health measures contribute to raising health standards globally. Increased movement of goods or people, however, also creates health risks arising from the dissemination of diseases or contaminated products.

Income, Inequality, and Health Higher income is associated with better health standards, both within and across countries. Cross-country studies find that aggregate health measures such as life expectancy are positively correlated with income per capita (see Pritchett and Summers 1996; Deaton 2004). The link between income and life expectancy appears to be strongest for low-income countries, but there also is more inequality in terms of health outcomes between these countries than among higher-income countries. Thus, although the potential health gains associated with higher incomes appear to be greatest in low-income countries, these countries are also where the impact of other factors on health outcomes is strongest, including (lack of) public health infrastructure or inefficiencies in the delivery of public health services, some of which, in turn, may also hold back improvements in income (Deaton 2004).

Rapid economic development is sometimes associated with an increase in income inequality, which may have implications for access to health services across income groups. Deaton (2003) points to the link between poverty and health outcomes, and finds little evidence for a role of inequality per se, however. Cross-country data on health care financing point to complex interactions between access to health care, poverty, and inequality. Financial sector development, and thus access to health insurance, is limited in many developing countries. Conversely, in developing countries the share of out-of-pocket expenditures in private health expenditure is higher, which implies that health risks are associated with higher risks to material living standards. For poor segments of the population, public health expenditure can compensate for the lack of health insurance, but the level of public health expenditures, their efficiency, and geographical access differ considerably across countries.

The correlation between increases in income and increases in life expectancy has motivated some studies trying to assess the contribution of each to improving living standards. Composite indexes, such as the Human Development Index of the United Nations Development Program, assign weights to measures of income, health, and possibly other indicators such as educational attainment. Some studies draw on microeconomic estimates of the valuation of mortality risks. Using this approach, Nordhaus (2002) suggests that the contribution of improved health to living standards has been of a magnitude similar to the contribution of improvements in gross domestic product (GDP) for the United States between 1900 and 1995, and Becker, Philipson, and Soares (2005) suggest that world

health inequality has been declining even if incomes have not converged.

Knowledge Flows and Health Technologies International knowledge flows have played an important role in falling mortality rates across countries. Deaton (2004) asserts that health improvements in developing countries in the 20th century "ultimately came from the globalization of knowledge, facilitated by local political, economic, and educational outcomes," including applications of the germ theory of disease, DDT spraying against malarial mosquitoes, the use of antibiotics against tuberculosis, and oral rehydration therapy.

The prominent role of knowledge flows points to the value of international cooperation to facilitate and speed up the dissemination of health technologies, and to assist in their implementation, which is one of the key mandates of the World Health Organization (WHO). Recent examples of internationally coordinated health programs include the Measles Initiative, which reportedly succeeded in reducing global deaths from measles by 60 percent (and by 75 percent in sub-Saharan Africa) between 1999 and 2005, largely through mass immunizations and improvements in care. Another major international health campaign was the WHO's 3 by 5 campaign and related efforts to reduce the incidence of HIV/AIDS and to improve access to antiretroviral treatment in low- and middle-income countries.

A key factor in the global dissemination of health technologies is the legal framework for the protection of intellectual property rights, including the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), which requires all World Trade Organization member countries to meet certain minimum standards for the protection of intellectual property rights, including patents. Patents allow pharmaceutical companies to recoup the costs of drug development by charging prices that far exceed production costs. Although most essential medicines (according to the WHO definition) are no longer protected by patents, patent protection can complicate access to medicines for relatively new diseases, such as HIV/AIDS. The experience with antiretroviral drugs to treat HIV/AIDS also involved numerous ways of facilitating access to these drugs in low-income countries, however.

Apart from its effect on the dissemination of medical know-how and drugs, the legal environment also has implications for the direction of research. As most of the profits from new drugs arise from sales in high-income countries, pharmaceutical companies direct their research efforts toward drugs for these markets, rather than drugs addressing health issues that are more prevalent in low-income countries.

Health, Growth, and the Location of Economic Activity While higher income can buy better health, the state of health can also have implications for income levels and affect the location of economic activities. Most directly, healthier people are also more productive. Fogel (1993), discussing the effects of nutrition on increased labor supply in Britain between 1800 and 1980, argues that improved nutrition contributed to an increase in the labor force participation rate, and that a higher intake of calories increased workers' productivity. Together, Fogel estimates that these two effects account for 30 percent of economic growth in Britain over the 19th and 20th centuries.

Conceptually related are attempts to estimate the impact of health on productivity, using data on aggregate macroeconomic outcomes and health indicators. Among the wave of empirical studies on sources of growth starting in the early 1990s, numerous authors have included some measure of health, usually the level of life expectancy. For example, Bloom, Canning, and Sevilla (2001) find that one additional year of life expectancy raises GDP per capita by 4 percent; and Cole and Neumayer (2006) discuss the link between various measures of morbidity and total factor productivity. These studies, however, are usually not able to clearly distinguish between level of life expectancy and growth rate effects and — in light of the large range of proxies for health status used — offer few insights into the nature of the link between health and growth.

In addition to its effects on productivity, health may affect economic development through its impact on the rate of population growth. The logic of this is already present in the basic Solow growth

model, where higher population growth translates into a lower capital/labor ratio and thus lower output per capita. Along these lines, Acemoglu, Johnson, and Robinson (2006), studying the impact of declining mortality in the early 20th century, find that this has resulted in an increase in population growth and a slow-down in the rate of growth of per capita income. This effect has also played a role in explaining the limited impact of HIV/AIDS on economic growth in countries severely affected by the epidemic, as higher mortality translates to an increase in the capital/labor ratio; some studies, similarly, have illustrated the positive economic impact of health crises such as the Black Death and 1918 influenza epidemic on the incomes of survivors.

Another link between health and growth that has been intensively researched in recent years is the impact of health on institutional development. Acemoglu, Johnson, and Robinson (2003) highlight the role health conditions played in European colonial expansion. Where health conditions were hostile, Europeans set up extractive regimes with minimal investments of human resources and in institutional development, which tended to contribute little to the economic growth of the colonies. In environments more inviting to European settlers, European colonists were more likely to set up more egalitarian political and economic institutions that contributed more to the colonies' economic growth. The point of this institutional approach is that differences in health conditions have a persistent influence on economic outcomes through their impact on institutional development.

Although the evidence on the links between income and health suggests that health would improve with economic development, the link between globalization, development, and health may not be positive in all cases or for all regions or people involved. Specifically, health, environmental, and safety standards are generally less stringent in developing countries, which translates into a cost advantage for companies producing there. In a world of competitive markets, the stringency of health and other standards related to health may reflect legitimate societal preferences regarding the importance of achieving certain health outcomes compared to other development objectives such as raising incomes or reducing poverty. Thus a government may accept lower health and environmental standards to attract investment and pursue its broader development strategy, and employees may accept health risks in return for higher salaries.

The most compelling arguments against trade based on differences in health and environmental standards across countries rest on asymmetric-information, incomplete-market, or moral arguments. The correct pricing of health risks requires that both sides understand the risks involved, a condition that often does not hold in employment relations in developing countries. Similarly, the absence of effective regulations or weak contract enforcement means that companies are not held accountable for negative externalities associated with their activities. For example, in the case of disposal of toxic waste, part of the costs of such processes may be literally "dumped" on the local population, with little or no accountability for the originator. From a moral perspective, certain production methods (frequently quoted examples are child labor and exports of toxic waste) may be regarded as morally questionable, which could motivate legal provisions in the source countries of exports, or increasingly common public campaigns by nongovernmental organizations targeting consumer behavior.

Health Risks Associated with the Movements of Goods and People One of the defining characteristics of globalization is an increasing intensity of trade in goods, which also has some implications for global health risks. Contaminated food and other supplies can amplify the spread of diseases globally, especially if long incubation periods are involved or the disease is new and the associated risks are not well understood. For example, one important channel of transmission of HIV in the early stages of the epidemic was the trade in contaminated blood products.

Most commonly, however, trade-related health risks are associated with contaminated food or other products for human consumption. One example is a spate of deaths in Panama in 2006 linked to drugs containing contaminated ingredients imported from

China. More frequently, health hazards associated with imported food arise from contamination with pesticides or industrial waste (e.g., in seafood) or bacteria such as salmonella.

Globalization complicates the task of ensuring food safety standards. First, national efforts to attain certain standards in food production need to be complemented by measures to ensure the quality of imported products, which could involve inspections of imported products at the port of entry, harmonization of standards, or measures to ensure that certain standards are met in the source countries (e.g., EU beef import requirements). At the same time, mass production of and increased trade in food and other products means that such products are sold over large regions and across borders. Effectively responding to a health emergency thus also requires the ability to trace products back to their origin and possibly recall affected products.

Globalization has also expanded the trade in illicit drugs. The United Nations Office on Drugs and Crime estimates that the value of the global market in illicit drugs was U.S. $90 billion in 2003, and that about 200 million people (5 percent of the world's population) consumed illicit drugs at least annually, and 110 million at least monthly. By far the most widely used drug is cannabis. The largest consumer market for illicit drugs is North America, accounting for about 44 percent of the world market, while the most important producing regions are Latin America for cocaine, and various Middle Eastern and Asian countries for opiates.

Apart from trade in goods, a second element of globalization is the increased intensity of the movement of people, including a massive increase in global air travel. Between 1950 and 2005, the number of air travelers increased by a factor of 60 (from 31 million to 1.9 billion), and the number of passenger kilometers multiplied by a factor of 123 (from 28 million to 3.4 billion). As a consequence, it is increasingly difficult to contain an infectious disease locally or regionally, an issue that has gained increased attention internationally following the SARS outbreak and in light of the risks attributed to avian flu. Similarly, the global spread of an epidemic such as the 1918 influenza would, under current conditions, occur much faster, allowing less time for preventive measures or for preparation of a health response, and might lead to bottlenecks in the availability of essential medicines.

See also access to medicines; globalization; global public goods; HIV/AIDS

FURTHER READING

Acemoglu, Daron, Simon Johnson, and James Robinson. 2003. "Disease and Development in Historical Per spective." *Journal of the European Economic Association* 1 (2 3): 397 405. An influential paper discussing the impact of disease on economic development in the long term.

. 2006. "Disease and Development: The Effect of Life Expectancy on Economic Growth." NBER Work ing Paper No. 12269. Cambridge, MA: National Bu reau of Economic Research. A discussion of the impact of increased life expectancy on growth in the early 20th century.

Becker, Gary S., Tomas J. Philipson, and Rodrigo R. Soares. 2005. "The Quantity and Quality of Life and the Evo lution of World Inequality." *American Economic Review* 95 (1): 277 91. Discusses evolution of world inequality in terms of incomes and health outcomes.

Bloom, David E., David Canning, and J. Sevilla. 2004. "The Effect of Health on Economic Growth: A Pro duction Function Approach." *World Development* 32: 1 13. Estimates the impact of health expenditures on economic growth.

Cole, Matthew A., and Eric Neumayer. 2006. "The Impact of Poor Health on Total Factor Productivity." *Journal of Development Studies* 42 (6): 918 38.

Deaton, Angus. 2003. "Health, Inequality, and Economic Development." *Journal of Economic Literature* 41: 113 58. Estimates the impact of poor health on worker productivity, controlling for differences in factors such as education levels and the use of capital.

. 2004. "Health in an Age of Globalization." In *Brookings Trade Forum 2004*, edited by Susan Collins and Carol Graham. Washington DC: The Brookings Institution, 83 130. Both works by Deaton are good starting points for further reading on health and glob alization.

Fogel, Robert W. 1993. "Economic Growth, Population Theory, and Physiology: The Bearing of Long Term Processes on the Making of Economic Policy." *American Economic Review* 84 (3): 369 95. Nobel Lecture, also available online at http://nobelprize.org/nobel prizes/ economics/laureates/1993/fogel lecture.html. Discusses the contributions of improved nutrition to health and economic growth in Britain over the 19th and 20th centuries.

López Casasnovas, Guillem, Berta Rivera, and Luis Currais, eds. 2005. *Health and Economic Growth Findings and Policy Implications*. Cambridge MA: MIT Press. A comprehensive resource about the affect of health on economic growth.

Nordhaus, William D. 2002. "The Health of Nations: The Contribution of Improved Health to Living Standards." NBER Working Paper No. 8818. Cambridge, MA: National Bureau of Economic Research. Compares contributions of health and growth to improving living standards.

Pritchett, Lant, and Lawrence H. Summers. 1996. "Wealth ier Is Healthier." *Journal of Human Resources* 31 (4): 841 68. Another good starting point for further reading on health and globalization.

MARKUS HAACKER

■ Heckscher-Ohlin model

The Heckscher-Ohlin (H-O) model, which originated in Heckscher (1919) and Ohlin (1933) and was formalized and given narrower interpretation by Samuelson (1948), differs from the Ricardian theory of comparative advantage in two key respects. First, the Ricardian theory assumes only one factor of production, which robs it of any ability to address the internal income distribution effects of international trade. In contrast, the H-O theory allows for two factors of production, which opens the door to internal income distribution effects of trade. Second, whereas the Ricardian theory relies on the differences in technology across countries as the source of international trade, the H-O theory assumes the existence of the same technology everywhere and relies on the international differences in factor endowments as the basis of trade.

Because the H-O model allows economists to analyze the income distribution effects and plausibly gives a central role to intercountry differences in factor endowment rather than technology, which diffuses relatively rapidly internationally, it has come to serve as the main workhorse of trade theorists. The model leads to the conclusion that each country exports the goods that use its relatively abundant factor more intensively and such exports lead to a rise in the real and relative return to the latter. Symmetrically, the country imports products using its scarce factor more intensively, which lowers the real and relative return to the latter.

Recently, increased wage inequality in the rich countries as measured by skilled-to-unskilled wage has brought this model further to the center of the policy debate. Those favoring protection over free trade argue that just as the H-O model predicts, trade liberalization by skilled-labor-abundant rich countries has led to the rise in skilled-to-unskilled wage. Pro-free-trade economists argue, however, that the real culprit behind the phenomenon is technological advances in skilled-labor-intensive industries, which has led to a shift in demand in favor of skilled labor.

The Setting and Principal Results The following will largely focus on the strict two-factor, two-good, and two-country version of the H-O model, as formalized by Samuelson (1948). This version is narrower than originally conceived by Ohlin (1933) but is now widely used by trade economists. To formally outline the key assumptions and structure of the model, call the countries Home Country (HC) and Foreign Country (FC), products Corn (C) and Shirts (S), and factors Land (T) and Labor (L). Assume constant returns to scale in the production of each good. This means doubling the use of each factor in C doubles the output of C. The same holds true for S. Goods C and S require different technologies. Specifically, we assume that C is land intensive in the sense that at any given set of factor prices, the land-labor ratio in the production of C is higher than that in S: $T_C/L_C > T_S/L_S$. Here

T_C denotes the quantity of land employed in the production of C. A similar interpretation applies to L_C, T_S, and L_S. By implication, S is relatively labor intensive. Assuming both goods are produced in equilibrium and there is perfect competition in all markets, the H-O model leads to the following two theorems:

The Stolper-Samuelson (1941) theorem: An exogenous increase in the relative price of a good leads to an increase in the real and relative return to the factor used more intensively in that good and a decrease in the real and relative return to the other factor. For example, an increase in the price of C increases the real and relative return to land, which is used more intensively in the latter. It also leads to a decline in the real and relative return to labor, which is used more intensively in S.

The Rybczynski (1955) theorem: Holding the goods prices constant, an increase in the endowment of a factor leads to a proportionately larger increase in the output of the good using that factor more intensively and a decline in the output of the other good. For example, holding the goods prices constant, a 1 percent increase in the endowment of labor would raise the output of S, which uses labor more intensively, by more than 1 percent and lower the output of C.

Suppose we additionally assume that HC is land abundant relative to FC. Formally, using an asterisk to distinguish the variables associated with FC, $T/L > T^*/L^*$. The two countries have the same technology of production available. These two additional results follow:

The Heckscher-Ohlin theorem: The opening to trade leads each country to export the good using its abundant factor more intensively and to import the other good. For example, the land-abundant HC exports the land-intensive good C and imports the labor-intensive good S.

The Factor-Price Equalization theorem: Absent transportation costs, the opening to trade equalizes not only the goods prices but factor prices as well. That is to say, in a trading equilibrium, the HC ends up with the same real and relative wage and rental price of land as FC.

Deriving the Theorems Let us denote the wage by w and the rental price of land by r. The proportion of land to labor in C, T_C/L_C, depends on w/r. Specifically, as labor becomes more expensive relative to land, firms economize on the use of labor and employ land more liberally. In other words, as w/r rises, they raise the land-labor ratio. This relationship is shown by curve CC′ in figure 1.

We assume that C is land intensive relative to S. This means that at each wage-rental ratio, C uses more land per unit of labor than S. In terms of figure 1, the line showing the land-labor ratio in S at different wage-rental ratios, SS′, lies everywhere below CC′.

Represent the land-labor ratio of HC in figure 1 by T/L. We can then determine the range of possible wage-rental ratios in HC. At a sufficiently high relative price of C, the economy specializes completely in this product. In this case, the land-labor ratio in C coincides with the economy's land-labor endowment ratio, and the wage-rental ratio is given by $(w/r)_{min}$ in figure 1. At the other extreme, if the price of C is sufficiently low, the economy specializes completely in S, the land-labor ratio in S coincides with the economy's land-labor ratio, and the wage-rental ratio is given by $(w/r)_{max}$ in figure 1. At an intermediate price ratio, both goods are produced and the wage-rental ratio is between the two extremes.
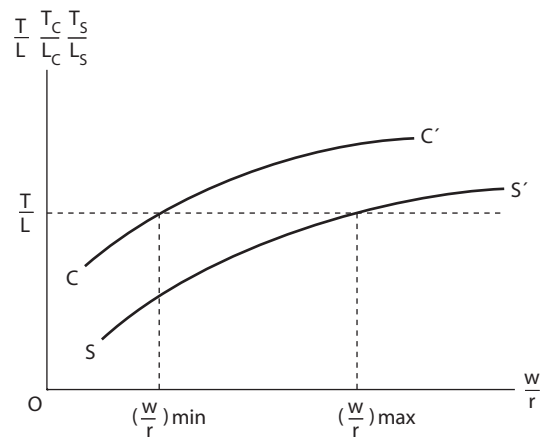
Figure 1
Factor prices and factor intensities

As we move from $(w/r)_{min}$ toward $(w/r)_{max}$, we increase the output of S and reduce that of C. With the wage-rental ratio rising during this movement, both sectors raise the land-labor ratio, and thus allow the labor-intensive sector S to expand.

Figure 2 shows the output changes just described using the construction of the production possibilities frontier (PPF). Given technology and factor endowments, MN shows the PPF of HC. The absolute value of the slope of the PPF at any point gives the opportunity cost of the product on the horizontal axis (C) in terms of the product on the vertical axis (S). The PPF is bowed out, which means that the marginal opportunity cost of each product in terms of the other rises as we expand the output of that product. For example, as we move from M toward N, the absolute value of the slope rises, meaning that the opportunity cost of C in terms of S rises with the rising output of C. This rising cost is the result of the imperfect substitutability between land and labor. To expand the production of C, which is land intensive, we must lower the land-labor ratio in each product. The more we substitute labor for land, the lower the marginal return to such substitution and the higher the marginal cost of further expansion.

Under perfect competition, production takes place at a point where the marginal cost equals the price.

Therefore, letting $(P_C/P_S)^0$ represent the relative price, the economy would produce at a point such as Q^0 in figure 2 where the price line is tangent to the PPF. The tangency ensures that the relative price equals the marginal opportunity cost of production.

Starting at Q^0, suppose we consider a small increase in the relative price of C. This would lead to an increase in the output of C and decrease in the output of S. At the original factor prices, the expansion of C would require more land per worker than S releases. This creates an excess demand for land and excess supply of labor. The return to land rises and that to labor falls. That is to say, a rising price of C, which is land intensive, is associated with a declining w/r ratio.

Let us plot these relationships in figure 3. In the right-hand panel, curve RS shows increasing supply of C relative to S as a function of the relative price of C, P_C/P_S. In the left-hand panel, curve RR' shows an inverse relationship between wage-rental ratio and the relative price of C. This latter relationship partially represents the Stolper-Samuelson theorem: it connects the goods prices to the relative factor prices. The representation is only partial since the Stolper-Samuelson theorem also relates the price change to the real factor returns.

To elaborate on how the Stolper-Samuelson theorem works, consider a move from autarky to free trade by HC. Suppose that the world relative price of C exceeds its autarky price. This means HC would export C and import S under free trade. As already

Figure 2
The production equilibrium
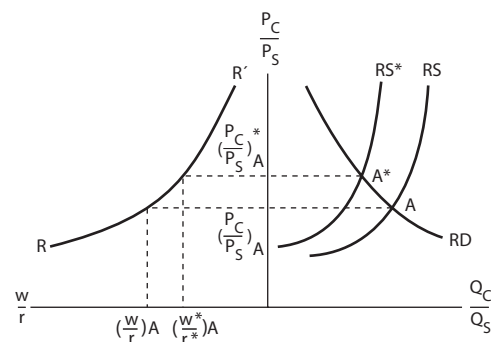
Figure 3
The Heckscher Ohlin factor price equalization theorems

explained, the higher price of C brought about by trade would increase the output of C and lower the wage-rental ratio (see left-hand panel in figure 3). The income distribution within HC would move against workers and in favor of landowners. The critical question is whether the workers could still be better off in real terms due to the decline in the relative price of good S brought about by trade. The Stolper-Samuelson theorem answers this question in the negative: the wage declines not just in terms of good C whose price rises but also in terms of S whose price falls.

To see how this works, note that the firms employ workers up to the point where the wage equals the value of marginal product of labor. Denoting the marginal product of labor in C and S by MPL_C and MPL_S, respectively, we have

$$w = P_C . MPL_C = P_S . MPL_S \qquad (1)$$

Rearranging, this equation implies

$$w/P_C = MPL_C \text{ and } w/P_S = MPL_S \qquad (2)$$

That is to say, the purchasing power of the wage in terms of a commodity equals the marginal product of labor in that commodity. This purchasing power rises or falls as the marginal product of labor in the product rises or falls. A similar relationship applies to land: the purchasing power of rental income in terms of a good rises or falls as the marginal product of land in the production of that good rises or falls.

To determine what happens to the marginal product, recall that as C expands, the wage-rental rate falls, which leads to a decline in the land-labor ratio in each product. The decline in the land-labor ratio implies a decline in the marginal product of labor and a rise in the marginal product of land in terms of each product. It follows that the real return to labor falls and that to land rises in terms of each good. In effect, the decline in the wage is sharper than the decline in the price of good S, leaving the workers worse off even if they spend their entire wage income on that good.

Next, let us consider the effects of a change in the factor endowments at a given goods price ratio. To take a concrete example, let us increase the endowment of labor by 5 units. From curve RR′ in figure 3, we know that as long as we hold the goods price

constant, the relative factor prices remain constant as well. Figure 1 then tells us that the land-labor ratios in the two goods must also remain unchanged. The economy must absorb the additional labor supply without altering the land-labor ratios. This constraint immediately rules out a simple-minded division of the additional units of labor between the two sectors without reallocation of land since such allocation would necessarily change the land-labor ratio in each product.

To see what kind of reallocations would be compatible with full employment at unchanging land-labor ratios, begin by placing the entire additional labor supply in the labor-intensive good S. To maintain the original land-labor ratio, this requires drawing land from C to work with the additional five units of labor in S. But since C must also maintain its original land-labor ratio, it would not release land without releasing labor. This means that S must absorb not just the 5 new units of labor but also those released by C as it releases land.

For concreteness, suppose the land-labor ratio is 3 in C and 2 in S. Then each time C releases 3 units of land, it releases 1 unit of labor. But since the land-labor ratio in S is 2, it employs 1.5 units of labor for each 3 units of land. In other words, moving 3 units of land out of C allows S to absorb 1 unit of labor released by C plus a half unit out of the new 5 units. Therefore, if we move 30 units of land from C to S, the latter would absorb 15 units of labor that are exactly equal to the sum of the 10 units released by C and 5 new units. Full employment is achieved at unchanging land-labor ratios.

This example illustrates that a given expansion of the endowment of a factor leads to a proportionately larger expansion of the sector using that factor more intensively and a contraction of the other sector. We thus have the Rybczynski theorem, named after T. M. Rybczynski (1955), who first noted the result. The result is generalized in the sense that if both factors expand but the land-labor ratio declines, the output of the land-intensive good relative to the labor-intensive good falls as well.

The Rybczynski theorem is a key building block of the Heckscher-Ohlin theorem. Recall that we

have assumed FC to be relatively labor abundant: $L^*/T^* > L/T$. An immediate implication of the Rybczynski theorem is that at any given price ratio, FC produces less C relative to S than HC. Therefore, denoting by Q_i the quantity of output of good i (i = C, S), we have $Q_C^*/Q_S^* < Q_C/Q_S$ at each price ratio. This is shown by the relative supply curve RS* in figure 3.

Assuming the relative demand depends only on the relative price and the consumers in HC and FC are identical, we can represent the demand in the two countries by a common demand curve RD in figure 3. The autarky equilibriums in HC and FC are then given by A and A*. It is straightforward that under autarky the land-intensive good C is cheaper in the land-abundant country HC and the labor-intensive good S is cheaper in the labor-abundant country FC. Therefore, when the two countries open to trade, each country would export the good that uses its abundant factor more intensively, just as the Heckscher-Ohlin theorem predicts (see above).

Finally, observe that, as the left-hand panel of figure 3 shows, under autarky the wage-rental rate is higher in the labor-scarce HC than in the labor-abundant FC. The opening to trade leads to a rise in the relative price of C in HC. This leads to a fall in the wage-rental rate there. The opposite happens in FC: the relative price of C declines there, which leads to a rise in the wage-rental rate. Therefore, the factor prices converge between the two countries. Assuming no transport costs, trade would equalize the goods prices. But as goods prices equalize, the wage-rental rate would equalize as well.

Given the same wage-rental rates in the two countries at the free-trade equilibrium, figure 1 tells us that the land-labor ratios across countries would also equalize for each product. This would then lead to the equalization of the marginal products and hence real factor prices of each product. Free trade would lead to the equalization of the relative and real factor returns internationally, as predicted by the factor price equalization theorem (see above).

Trade and Wages In the contemporary policy literature, the Stolper-Samuelson and factor price equalization theorems have played a crucial role.

Between the late 1970s and early 1990s, the real and relative wages of unskilled workers in relation to skilled workers in the rich countries declined. The ratio of skilled-to-unskilled wages, a measure of wage inequality, rose almost 30 percent in the United States. This period also coincided with a rapid expansion of trade between developed and developing countries. This led many to link the changes in the wages to the opening to trade via the Stolper-Samuelson theorem. If we think of the two factors in the H-O model as skilled and unskilled workers, the developed countries are importers of unskilled-labor-intensive goods. The H-O theory then predicts that opening to trade with the developing countries would push down the real and relative wages of the unskilled. In the spirit of the factor price equalization theorem, some observers have gone so far as to suggest that the wages of the unskilled in the developed countries may be pushed down to the levels prevailing in the developing countries.

Trade economists disagree with this diagnosis and argue that trade with the poor countries cannot explain the bulk of the increase in the wage inequality. They cite four reasons in support of their position.

- Trade works to lower the unskilled wage by lowering the relative price of unskilled-labor-intensive goods. But a study by Lawrence and Slaughter (1993) pointed out that the relative price of unskilled-labor-intensive goods had actually risen since the late 1970s. By itself this point is not decisive, however. In principle, trade may have lowered the relative price of unskilled-labor-intensive products but other factors such as sharply declining costs of skilled-labor-intensive products may have reversed this decline. Lawrence and Slaughter only looked at the ex post change in the prices but did not decompose them according to the sources of the change. Therefore, their analysis remains incomplete.

- Extra imports from the developing countries during the relevant period account for less than 2 percent of the total expenditure

in the United States. This is the point made by Krugman (1995). He argues that such a small proportionate expansion of trade can simply not explain the large increase in wage inequality. This is a valid and important point.

- During this period, wage inequality rose in many developing countries as well. If the simple-minded Stolper-Samuelson theorem was driving the outcome, developing countries should have experienced a decrease in wage inequality. Given that they export unskilled-labor-intensive goods, the Stolper-Samuelson theorem should have driven their real and relative unskilled wages up. This did not happen.

- Technical change that shifted labor demand in favor of skilled labor and away from unskilled labor provides a far more compelling explanation for increased wage inequality in both rich and poor countries. Technological change has been concentrated in skilled-labor-intensive goods, and it has also moved progressively toward greater use of skilled labor. This change has shifted the demand in favor of skilled labor in both rich and poor countries and led to increased wage inequality in both regions.

The Heckscher-Ohlin model is the principal workhorse of trade economists. It shows that countries export goods that use their abundant factors more intensively and import goods that use their scarce factors more intensively. Given that imports bring goods that use the scarce factors more intensively, they lower the demand for and hence the returns to such factors locally. This conclusion has led to widespread claims that the imports of unskilled-labor-intensive goods from the poor countries have lowered real and relative wages of unskilled workers in the rich countries. Most trade economists disagree with this conclusion, arguing that a shift in technology in favor of skilled labor and away from unskilled labor is the true cause of the decline in the fortunes of unskilled workers.

See also comparative advantage; factor endowments and foreign direct investment; Ricardian model; specific-factors model; trade and wages

FURTHER READING

Bhagwati, Jagdish. 1964. "The Pure Theory of International Trade: A Survey." *Economic Journal* 74 (293): 1 84. Offers a careful survey of the development of the Heckscher Ohlin theory along with other theories of international trade.

Heckscher, Eli. 1919. "The Effect of Foreign Trade on the Distribution of Income." *Ekonomisk Tidskrift* 21: 497 512. The original article to which the Heckscher Ohlin theory is traced.

Krugman, Paul. 1995. "Growing World Trade: Causes and Consequences." *Brookings Papers on Economic Activity* 1: 327 62. Argues that the rise in volume of imports from the developing countries in relation to the total expen diture is too tiny to explain the large increase in wage inequality during the 1980s and early 1990s in the United States.

Lawrence, Robert Z., and Mathew J. Slaughter. 1993. "International Trade and American Wages in the 1980s: Giant Sucking Sound or a Small Hiccup?" *Brookings Papers on Economic Activity* 2 (Microeconomics): 161 226. Shows that the prices of unskilled labor intensive goods relative to skilled labor intensive goods rose dur ing the phase that the unskilled wages fell in the United States.

Ohlin, Bertil. 1933. *Interregional and International Trade*. Cambridge, MA: Harvard University Press. First full scale statement of the Heckscher Ohlin theory in English by one of its two originators.

Panagariya, Arvind. 2000. "Evaluating the Factor Content Approach to Measuring the Effect of Trade on Wage Inequality." *Journal of International Economics* 50 (1): 91 116. Offers a detailed analysis of the relationship of the factor content of trade to factor prices in the H O model.

Rybczynski, T. M. 1955. "Factor Endowments and Relative Commodity Prices." *Economica* 22 (87): 336 41. Inaugurated the systematic analysis of the effects of the changes in factor endowments on outputs.

Samuelson, Paul A. 1948. "International Trade and the Equalization of Factor Prices." *Economic Journal* 58

(230): 163 84. Formalizes the Heckscher Ohlin model and shows that trade would lead to the equalization of factor prices.

Stolper, Wolfgang, and Paul A. Samuelson. 1941. "Pro tection and Real Wages." *Review of Economic Studies* 9 (3): 58 73. Shows that in the H O model, the owners of the scarce factor are left unambiguously worse off by imports.

ARVIND PANAGARIYA

■ hedge funds

Hedge funds are broadly defined as private invest ment pools that are not available to the general public. They are more lightly regulated and have wider investment flexibility than public investment companies such as mutual funds, which pool money from many investors and invest in stocks, bonds, and other securities. Hedge funds can buy and sell secu rities in many financial markets, representing long and short positions, respectively. In addition, they can use leverage and derivatives, which are financial instruments whose value derives from some under lying asset or price.

The first hedge fund was started in 1949 by Alfred Winslow Jones, a financial journalist, who believed that this new investment style could deliver good returns with more stability than investments in stock mutual funds. The hedge fund industry has grown from 600 funds in 1990 to more than 8,500 in 2005. During the same period, total assets in hedge funds have grown from $40 billion to more than $1 trillion. This growth has been driven by the stable investment performance of the industry, especially when com pared to the swings of the stock market. Because they pursue very active investment strategies, hedge funds are even more important than their asset size would suggest. As a result, hedge funds have become major players in international capital markets.

Types of Hedge Funds The hedge fund industry is much more heterogeneous than the mutual fund industry because of the greater latitude in investment style. Funds are typically classified into the following categories:

Global macro funds, which take positions in global markets (stocks, fixed-income in vestments, currencies, commodities);

Long/short equity funds, which buy and sell stocks;

Equity market neutral funds, where the long positions are exactly offset by short posi tions so as to create a zero, or neutral ex posure to the stock market;

Arbitrage funds, which take long and short positions in securities such as fixed-income and convertible bonds;

Event-driven funds (merger arbitrage and dis tressed debt), which take positions driven by corporate events such as mergers, takeovers, reorganizations, and bankruptcies.

Long/short equity funds represent the largest sector of the industry, with approximately one-third of the funds.

Investment Strategy Consider a typical hedge fund, which has both long and short positions in stocks. Say the initial capital is $100. This represents the equity, or net asset value. The fund buys $100 worth of stocks and sells $50 worth of other stocks. Short-selling is achieved by borrowing a stock and selling it in the hope that its price will fall later, at which time the stock can be bought back and trans ferred to the lender. In such case, the borrower keeps the difference between the (higher) earlier sales price for the stock and the (lower) later purchase price. In the event that the stock price goes up, however, the borrower loses the difference between the (lower) ear lier sales price and the (higher) later purchase price.

This type of investment strategy (short-selling) has two advantages relative to mutual funds, which typically are allowed to have long positions only. First, it allows the hedge fund manager the flexibility to buy assets that are viewed as undervalued, for ex ample, and sell assets that are overvalued. In contrast, the manager of a long-only fund cannot implement a view that an overvalued asset is going to fall in price, because the manager cannot short the asset.

Second, it has less exposure to the direction of the stock market (called *directional exposure*) than a long- only position. Indeed, to "hedge" a bet can be defined

as protecting against loss by taking a bet for a countervailing amount against the original bet. In hedge investing, the directional exposure to the market, also called *beta*, is reduced. As a result, the overall risk of loss is reduced as well. Hedge funds attempt to create risk-adjusted performance, also called *alpha*, without taking too much directional risk.

As an example, a manager could buy Google's stock because it is expected to go up by 25 percent, out of which 10 percent is due to the overall stock market and 15 percent to the company itself. The manager, however, is not confident about the direction of the stock market and is afraid the 10 percent expected profit on the market could turn into a loss. To hedge against a market fall, the purchase of Google is offset by a short position of the same size on the market. If the market falls by 10 percent, the transaction will create a profit of 15 percent minus 10 percent, for a 5 percent total gain on Google's stock, plus a gain of 10 percent on the stock market, for a total of 15 percent. Of course, if the market indeed goes up by 10 percent, the short position will lose 10 percent, leading to a net profit of 15 percent. Therefore, this hedge locks in a total profit of 15 percent irrespective of the market.

Another example of trading strategy that exploits inefficiencies in the market would be to identify two assets with similar characteristics, such as two long-term Treasury bonds with close maturities and similar interest rate risk. An *arbitrage* opportunity would arise if the bonds were to trade at sufficiently different prices. The hedge fund manager would then buy the cheap bond, say at $99, and sell the expensive one, say at $100. This position should create profits if the bond prices later converge. Even if successful, however, such strategy will produce only small profits, on the order of $1 per transaction (after taking into account the costs of making the transaction). To magnify profits, the hedge fund will typically use *leverage*. This involves the use of credit to increase the size of the position relative to the equity. For instance, the hedge fund with $100 in equity could be buying and selling $990 and $1,000 worth of bonds (using 10 percent equity and 90 percent credit), which will multiply the profit by a factor of 10.

This wider investment flexibility combined with hedging has proved successful. Hedge funds have generally performed well relative to other investments, especially when adjusting for risk. For fund managers, hedge funds provide greater remuneration than traditional investment funds. Typical investment management fees for mutual funds range from a fixed 0.5 percent to 2 percent of asset value. In contrast, hedge funds commonly charge a fixed 2 percent of assets plus 20 percent of positive returns.

Other categories of hedge funds include *multistrategy funds*, which invest across the categories listed earlier, and *funds of funds*, which invest in individual hedge funds.

Investment Issues Hedge funds take risks in the expectation of high profits. Because of the leverage inherent in most hedge funds, this naturally leads to failures. Leverage increases the profits, but also the risks. In our previous example, a leverage of 10 transforms a profit of $1 into $10. On the other hand, a loss of $1 now becomes $10. In the limit, a very large loss of $10 becomes $100, wiping out the $100 equity. Indeed, hedge funds have a high rate of disappearance, from 5 to 11 percent per year. In most cases, this reflects an investment or operational failure that leads to the closing down of the fund. Mutual funds also fail, but less frequently.

Hedge funds present some specific problems not shared by mutual funds. Because hedge funds often invest in assets that are not traded on organized exchanges such as the New York Stock Exchange, some positions can be difficult to value. Thus the proper *valuation* of assets is a major issue for hedge funds. Assets that are not traded often are called illiquid.

Illiquidity explains the common practice of imposing a lockup period, which is the minimum amount of time during which investor money is to be held in the fund, and a redemption notice period, which is the time required to notify the fund of an intended redemption. Lockup periods average 3 months, and can extend to 5 years; redemption notice periods average 30 days.

Another problem with hedge funds is the *lack of transparency*. Because hedge funds follow proprietary trading strategies, they are generally reluctant to re-

veal information about their trading ideas or positions. Having no information about positions has serious disadvantages for investors. It becomes difficult to monitor the risk of the fund, to aggregate the risk profile of the fund with the rest of the investor's portfolio, or even to detect fraud.

Regulatory Issues The rapid growth of hedge funds has led to regulatory concerns on several issues. One is the *protection of investors*. Investors can lose money in hedge funds. Direct investment in hedge funds, however, is restricted to professional investors, such as wealthy individuals, pension funds, endowment funds, and other institutional investors. The general presumption is that such investors are conscious of the investment risk they take and can therefore fend for themselves. Also, hedge fund managers generally have significant personal investments in the fund, which partially aligns their interest with that of other investors.

Fraud is another concern. Fraud occasionally occurs in all investment activities, and there is no evidence that hedge fund advisors engage disproportionately in fraudulent activities. Even so, there have been some high-profile cases of hedge fund fraud which, combined with the rapid growth of the industry, have led regulators to look more closely at hedge funds.

Another issue is *protecting market integrity*. Market participants should not attempt to dominate or manipulate markets. Like other large investors, hedge funds are subject to rules against market manipulation.

Regulators also worry about the potential for *systemic risk*. Systemic risk arises when default by one institution has a cascading effect on other financial institutions (the first institution's default being of such a magnitude as to cause the failure of its creditors, which in turn default on their obligations to other institutions, and so on), thus posing a threat to the stability of the entire financial system. Indeed, the spectacular failure of the hedge fund Long-Term Capital Management (LTCM) in 1998 is said to have endangered the world financial system. LTCM was a hedge fund founded in 1994 by John Meriwether, the former head of bond trading at Salomon Brothers. The fund lost $4.4 billion in 1998, forcing the Federal Reserve Bank to organize a private bailout to avoid disruptions in financial markets.

The primary mechanism for regulating hedge fund risk is the discipline provided by creditors, counterparties, and investors. The leverage of the hedge fund industry is provided by counterparties such as banks and brokerage firms. It is in these lenders' best interest to make sure that loans to hedge funds are controlled and have sufficient collateral. Holding collateral can provide effective protection against default. The LTCM crisis revealed lax lending standards. Since then, however, the Counterparty Risk Management Policy Group (2005) has reported a general improvement in counterparty risk management practices.

Role of Hedge Funds Hedge funds have become important conduits of capital in international financial markets. Like any other speculator trading in the expectation of making profits, hedge funds enhance market liquidity by generating more trading activity, leading to deeper and more liquid markets. In addition, many hedge funds use relative value strategies, which buy underpriced assets and sell overpriced ones. These actions should help to push prices faster to their equilibrium values, which enhance market efficiency.

See also capital mobility; carry trade; hedging; speculation

FURTHER READING

Chan, Nicholas, Mila Getmansky, Shane Haas, and Andrew Lo. 2006. "Systemic Risk and Hedge Funds." In *The Risk of Financial Institutions*, edited by Rene Stulz and Mark Carey. Chicago, IL: University of Chicago Press, 235 330. A review of risk measurement issues and fail ure rates in the hedge fund industry.

Counterparty Risk Management Policy Group. 2005. *To ward Greater Financial Stability: A Private Sector Per spective*. New York: CRMPG. An evaluation of potential systemic risk problems in financial markets.

Edwards, Franklin. 1999. "Hedge Funds and the Collapse of Long Term Capital Management." *Journal of Eco nomic Perspectives* 13 (spring): 189 210.

Eichengreen, Barry, Donald Mathieson, Bankim Chadha, Anne Jansen, Laura Kodres, and Sunil Sharma. 1998.

Hedge Funds and Financial Market Dynamics. Wash ington, DC: International Monetary Fund. A review of the hedge fund industry, its role in the Asian currency crisis, and regulatory issues.

Financial Services Authority. 2006. *Hedge Funds: A Dis cussion of Risk and Regulatory Engagement*. London: FSA. Perspectives from a major regulator of hedge funds based in the United Kingdom.

Gupta, Anurag, and Bing Liang. 2005. "Do Hedge Funds Have Enough Capital? A Value at Risk Approach." *Journal of Financial Economics* 77 (July): 219 53. A detailed analysis of hedge fund risks.

Jorion, Philippe. 2000. "Risk Management Lessons from Long Term Capital Management." *European Financial Management* 6: 277 300. The role of leverage and risk in the LTCM crisis.

PHILIPPE JORION

■ hedging

Hedging refers to the reduction of an existent risk by the elimination of exposure to price movements in an asset. Hedging stabilizes markets since it removes potential shocks to balance sheets that can destabilize the financial system. Also, if hedging is complete at the aggregate level, long and short positions can be matched with less price volatility. Those taking a short position want to sell an asset because they be lieve its price is going to fall, while those going long want to buy it since they expect a rise in the asset price. Unhedged short-term foreign borrowing played a major role in escalating the East Asian crisis of 1997 98.

Currency risk is hedged through contracts that protect the home currency value of transactions de nominated in a foreign currency, removing the ex posure to exchange rate fluctuations. The currency risk is transferred to another party who wants to take an exposure in opposite direction.

The value of a derivative hedging contract is based on the underlying basic spot exchange rate. Stan dardized contracts are available, or customized (over-the-counter, or OTC) contracts can be designed. An Indian exporter, for example, can sell the dollars due to him in the future on the dollar-rupee forward market, through a forward contract or an agreement to sell the dollars at a certain future rate for a certain price reflecting current prices. Hedging can also be accomplished informally, for example, if an exporter takes a foreign loan. Then if the home currency de preciates he will gain in export income but lose as the home currency value of his debt rises, so that his net exposure is low.

Hedging versus Speculation If the exporter does not sell the future dollars forward, thus entering into only one leg of a foreign currency transaction, he is speculating on a belief the rupee is going to depre ciate. He may also take a short position on the rupee using derivative products. Speculation is the act of aiming to profit by betting on a predicted one-way price movement. Thus it is risk-taking, not risk-re ducing. It is sometimes argued that since speculators buy when prices are low and sell when prices are high, rational speculative activity stabilizes markets. But this does not always follow, since speculators buy when there is a high probability of price appreciation and sell when the probability is low, and thus can cause cumulative movements.

Buying and selling currency in the spot market requires an initial cash payment. Forward contracts require no such initial payment, but can lead to a gain or a loss in the future transaction depending on changes in the exchange rate. Therefore, forward markets provide extra leverage to a speculator in comparison to spot markets. Options, another type of derivative, increase the leverage further since they confer a right but not an obligation to make a transaction. The option has an initial cost but then does not require buying or selling currency either in spot or in the forward market. Therefore it provides even higher leverage, at a price.

Incentives to Hedge Hedging does not neces sarily rise with the availability of more market in struments, since the same derivative can be used for hedging or for speculation leveraging initial capital many times. Incentives to induce hedging are more important. Markets, instruments, and opportunities,

with some restrictions on currency derivatives, existed in East Asia prior to the crisis, yet hedging was inadequate (Burnside et al. 2001).

Speculation dominates if there are aspects of the financial structure that encourage taking too much risk, such as lowering liability for promoters or managers; or if policy, such as one-way movement of exchange rates, creates opportunities for speculation. A firm's cost of borrowing is minimized if it is fully hedged so that bankruptcy does not occur. Since creditors normally set interest rates to cover expected default under bankruptcy, the risk-free interest rate is lower. So a fully hedged firm should get loans at lower rates. But if government guarantees the loan repayments and covers bankruptcy costs the incentive to hedge risks disappears (Burnside et al. 2001). Results are similar if laws make it difficult for creditors to recover loans.

If domestic interest rates are high, the high borrowing costs raise the opportunity costs of buying hedging instruments. Even if an importer holds a dollar deposit as an informal hedge, he sacrifices potential high domestic interest earnings. Similar considerations affect the hedging activities of banks.

Regulatory authorities sometimes put restrictions on derivative products that could facilitate hedging, in order to curb speculation. But with modern technology and regulation it is possible to distinguish between hedging and speculative foreign exchange market activities. Hedging transactions are charged lower margins on modern exchanges and their tax treatment also differs. If it is possible to monitor the use of derivatives it is not necessary to ban them.

Psychological factors also undermine rational hedging decisions. People prefer a sure gain, but they prefer an uncertain outcome with a small probability of a gain to a sure loss. Hedging involves a small sure loss or cost, and without it there is a small probability of a gain. Thus they are willing to forgo hedging and undertake more risk than is rational. But proper "framing" sensitive to psychological attitudes can reduce risk-taking behavior. An understanding that concentration on core business makes them more competitive and reduces volatility of profits may make firms more willing hedgers.

See also balance sheet approach/effects; currency crisis; financial crisis; hedge funds; speculation

FURTHER READING

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo. 2001. "Hedging and Financial Fragility in Fixed Ex change Rate Regimes." *European Economic Review* 45: 1151 93. Discusses the links between financial crises and failure to hedge.

Shefrin, Hersh. 2002. *Beyond Greed and Fear: Under standing Behavioral Finance and the Psychology of Investing*. Oxford: Oxford University Press. Applies psychology to analyze investor behavior.

Wang, Peijie. 2005. *The Economics of Foreign Exchange and Global Finance*. Berlin: Springer. A textbook treatment of hedging in currency markets.

ASHIMA GOYAL

■ HIV/AIDS

HIV/AIDS refers to the human immunodeficiency virus and the associated acquired immunodeficiency syndrome, which were first recognized as a global health issue in the early 1980s. In 2006, about 40 million people globally were living with HIV/AIDS, including 25 million in sub-Saharan Africa, and the epidemic claimed almost 3 million lives in that year. The international response has involved a rapid scaling-up of aid and an expansion of prevention and treatment programs, as well as the establishment of two specialized international agencies to deal with the epidemic the Joint United Nations Program on HIV/AIDS (UNAIDS) and the Global Fund to Fight HIV/AIDS, Tuberculosis, and Malaria (GFATM).

Epidemiology and Demographics The most important modes of transmission of HIV are sexual contact with an infected person, sharing needles for injected drug use, mother-to-child transmission before or during birth or through breast-feeding, and transfusions of infected blood or blood clotting factors. HIV/AIDS is largely asymptomatic during the

first years after infection. Over time, HIV/AIDS progressively damages the immune system, resulting in increased susceptibility to opportunistic infections (e.g., pneumonia, tuberculosis, and certain types of cancer) and, eventually, death. Antiretroviral treatment suppresses the virus and slows down the progression of the disease.

The region most affected by HIV/AIDS is sub-Saharan Africa; however, the high HIV prevalence rate of about 6 percent for the region masks very substantial differences in HIV prevalence across countries, ranging from less than 1 percent (e.g., in Senegal) to more than 20 percent in Botswana, Lesotho, Swaziland, and Zimbabwe (all data on HIV prevalence relate to the population ages 15 49, at end-2005). Although HIV prevalence appears to have stabilized in sub-Saharan Africa since around 2001, it has been spreading rapidly in Eastern Europe and Central Asia. Outside Africa, the country with the largest population of people living with HIV/AIDS is India.

Development Impact From an economic development perspective, the most direct impact of HIV/AIDS is the increase in mortality and consequent decline in life expectancy associated with it. For some of the most affected countries, it is estimated that HIV/AIDS has reduced life expectancy by more than 20 years. Development indexes such as the United Nations Development Program's Human Development Index, which combines measures of income, education, and health, suggest that HIV/AIDS has been the most significant single factor to adversely affect development in recent decades. Another consequence of the increase in mortality among young adults associated with HIV/AIDS is an increase in orphan rates, which are estimated to have reached 20 percent of the young population in some of the countries most affected by the epidemic.

The evidence regarding the macroeconomic impacts of HIV/AIDS is less clear. In light of the slowdown of the growth of the working-age population, there is a consensus that growth of gross domestic product (GDP) also slows as a consequence of HIV/AIDS. Studies applying a neoclassical growth model typically find that HIV/AIDS reduces GDP per capita through declining productivity and a fall in capital per worker that occurs as health-related spending lowers the national saving and investment rates. These effects are at least partly offset by the impact of increased mortality on the capital/labor ratio (mortality decreases the denominator, causing an increase in the overall ratio). The latter effect, however, partly dissipates if investment flows are sensitive to changes in the rate of return to capital.

On the microeconomic level, HIV/AIDS is associated with income losses (as household members become too sick to work and as working time is devoted to caregiving) and with increased health-related expenditures. Microeconomic data, largely from sub-Saharan Africa, therefore suggest an adverse impact of HIV/AIDS on incomes, consumption, and wealth of affected households. This is most pronounced during illness or around the time of death; households appear to partly recover later on. To understand the impact of HIV/AIDS on poverty or inequality, it is necessary to take a broader perspective, also covering households that may gain financially as they benefit from income opportunities associated with deaths in other households, most obviously when household members fill HIV/AIDS-related vacancies. Overall, the sparse evidence suggests that an increased volatility in incomes associated with higher mortality translates into an increase in poverty.

International Response On a global level, recognition of the health, humanitarian, and development challenges posed by HIV/AIDS has translated into an unprecedented effort to contain the epidemic and expand access to treatment. Important steps of the international response were the establishment of UNAIDS and the GFATM, and the United Nations General Assembly Special Session on HIV/AIDS (2004). In financial terms, the scale of the international response to HIV/AIDS has expanded very rapidly. Consistent estimates for HIV/AIDS-related spending are available for low- and middle-income countries only; for these, spending has increased from about U.S. $300 million in 1996 to about U.S. $9 billion in 2006, of which about U.S. $6 billion

was financed by external aid. The most important funding agencies are the GFATM and the U.S. President's Emergency Plan for AIDS Relief.

National responses to HIV/AIDS depend on the state of the epidemic in a country. To various extents programs emphasize public prevention measures in schools and work places, prevention and awareness measures targeted at high-risk groups, strengthening the health care system, improvements in care for people living with HIV/AIDS, measures to mitigate the social impacts (including support for orphans), and programs to expand access to treatment.

The most effective prevention measures are those targeted at groups at high risk of contracting and passing on the virus, including promotion of condom use among sex workers and provision of sterile needles to injecting drug users. Those measures, together with the perceived impact of the epidemic, have been credited with increasing awareness and reducing risky behavior. Social attitudes, particularly toward men who have sex with men, or the illegal nature of some of the risky behavior such as injecting drug use, can complicate the implementation of prevention programs, however.

In the most affected countries in Southern Africa, the epidemic is generalized, and prevention efforts are geared toward raising awareness and reducing risk behavior across the population, especially among young adults, through the education system, media and advertising campaigns, and public endorsements by leading politicians. Although HIV prevalence has risen to double-digit levels in numerous countries in spite of these efforts, in 2005 6 many of the most affected countries reported increasing HIV awareness and somewhat falling prevalence rates among young adults.

The most significant development in the early years of the 21st century regarding the response to HIV/AIDS was the decline in the costs of antiretroviral treatment. In many developing countries, certain forms of antiretroviral treatment were available in 2007 at costs of around U.S. $300 annually, down from about U.S. $10,000 in 2000. This development reflected voluntary agreements with drug companies, often under the threat of compulsory

licensing to a local producer, and the fact that only a certain range of antiretroviral drugs was available at these low prices, which allowed for some market segmentation between industrialized and developing countries.

Falling prices of drugs and strong international financial support have contributed to a rapid expansion in access to treatment. UNAIDS reports that the number of people receiving antiretroviral treatment in low- and middle-income countries increased from 400,000 to 1.3 million between 2003 and 2005 (corresponding to a coverage rate of about 20 percent), with sub-Saharan Africa accounting for the bulk of the increase.

Continued spread of the disease and the longer survival of those already infected make the management of an increasing number of people requiring treatment the principal challenge in addressing the epidemic in the near future. Additional challenges include extending the gains made to countries with weaker public health systems, where progress in expanding access to treatment has been less pronounced so far, and managing the fiscal challenges and long-term commitments associated with the expansion in these health programs.

UNAIDS and the Global Fund The perception of HIV/AIDS as a threat to global health, beyond the capacity and expertise of any single international organization, resulted in the establishment of a unique institution, UNAIDS, in 1994. UNAIDS coordinates the HIV/AIDS-related activities of its cosponsoring organizations, which are 10 (initially 6) organizations under the UN system. Although a relatively small organization on an international scale (its annual operational budget amounted to about U.S. $58 million in 2006 7), it is also financing part of the activities of its cosponsors on HIV/AIDS, as well as interagency activities (U.S. $42 million annually in 2006 7). Moreover, cosponsors include all of their HIV/AIDS-related activities in UNAIDS's Unified Workplan, which brings the total of HIV/AIDS-related spending coordinated by UNAIDS to about U.S. $1.3 billion annually for 2006 7, about one-sixth of global spending on HIV/AIDS. Additionally, UNAIDS is a key provider of public

information on the epidemic, including the annual Report on the Global AIDS Epidemic, which is the most important regular publication on HIV/AIDS.

The GFATM is primarily a funding agency. It receives about 95 percent of its funding from government sources. Grants typically underwrite comprehensive country programs, which are coordinated nationally, for several years. Between 2002 and end-2006, the GFATM disbursed U.S. $3.3 billion, of which U.S. $1.35 billion was disbursed in 2006. Of the accumulated grant portfolio, HIV/AIDS accounts for the lion's share (56 percent), followed by malaria (27 percent) and tuberculosis (15 percent). Although public institutions play the most important role as implementing agencies (accounting for about half of GFATM-supported funding), many of the national responses are implemented by nongovernmental organizations (about a quarter of funding) or by faith-based and academic organizations. Reflecting the burden of HIV/AIDS in the region, sub-Saharan Africa accounts for more than half of GFATM funding.

See also access to medicines; global public goods; globalization

FURTHER READING
Many key references are updated periodically and available online. Useful web sites are the ones of UNAIDS (www.unaids.org), including the annual Report on the Global AIDS Epidemic; the GFATM (www.theglobalfund.org); and the Global AIDS Program of the U.S. Centers for Disease Control and Prevention (www.cdc.gov/nchstp/od/gap/). The World Bank's website also includes much useful material, accessible from http://www.worldbank.org/aids. Other refer ences include:

Beck, Eduard J., Nicholas Mays, Alan W. Whiteside, and Jose M. Zuniga, eds. 2006. *The HIV Pandemic Local and Global Implications*. Oxford: Oxford University Press.

Canning, David. 2006. "The Economics of HIV/AIDS in Low Income Countries." *Journal of Economic Per spectives* 20 (3): 121 42. An overview of key economic issues in the treatment and spread of HIV/AIDS in poor countries.

Haacker, Markus, ed. 2004. *The Macroeconomics of HIV/AIDS*. Washington, DC: International Monetary Fund. Available online at http://www.imf.org/external/pubs/ft/aids/eng/index.htm. A collection of essays addressing economic, social, and demographic challenges to eco nomic development raised by the AIDS epidemic.

Jamison, Dean T., Richard G. Feachem, Malegapuru W. Makgoba, Eduard R. Bos, Florence K. Baingana, Karen J. Hofman, and Khama O. Rogo, eds. 2006. *Disease and Mortality in Sub Saharan Africa*. Washington, DC: World Bank. Places the AIDS epidemic in the broader context of African health and disease.

MARKUS HAACKER

■ home country bias

Home country bias, or simply *home bias*, usually refers to a situation in which the proportion of foreign equities held by domestic investors in their portfolios is too small relative to the predictions of standard portfolio theory. The extent to which equity portfolios are concentrated in equities of the investor's domestic market is a notable feature of international portfolio investment and has remained an important yet unresolved empirical puzzle in financial economics since the 1970s (Levy and Sarnat 1970). Since portfolio theory is the foundation of asset pricing theory, the empirical evidence that investors may not optimize along objective risk-return trade-offs as portfolio theory predicts has important implications for our understanding of the way security prices are set.

Standard models of optimal international portfolio diversification imply that equity investors have not diversified internationally nearly as much as they should. If investors only care about the mean and the variance of the real return of their portfolio, one would expect investors, as a first approximation, to hold the world market portfolio of stocks. In empirical finance literature, home bias is essentially measured by the underweighting of foreign securities in portfolio allocation relative to market-capitaliza-

tion weights. For example, by the end of 1989, Americans held roughly 94 percent of their equity portfolio in the U.S. stock market, even though the U.S. equity market comprises less than 48 percent of the global equity market (French and Poterba 1991). By the mid-1990s only about 10 percent of U.S. equity wealth was invested abroad, still much less than what it would have been had the U.S. investors held the world market portfolio (Tesar and Werner 1998). This phenomenon exists in other countries as well. A study based on worldwide equity fund holdings data in 1999 and 2000 documents the existence of home bias in every single country in its sample of 48 countries across the globe (Chan, Covrig, and Ng 2005). The literature has also shown that home bias exists not only among individual investors but also among institutional investors.

At the country level, the benchmark for measuring home bias is the share of equity of that country in the world market portfolio. For example, U.S. equity market capitalization accounts for about 50 percent of the world, and therefore the benchmark for U.S. portfolio diversification is 50 percent. At the individual level, the specific optimal mix depends on an individual's utility function, which should include factors such as risk aversion and investment horizon. Foreign holdings of all U.S. based mutual funds (who are institutional investors) that invest in international stocks are considered foreign investment in the literature.

Possible Explanations of Home Country Bias
International financial markets offer a means for diversification across markets. Most investors do not fully exploit this risk-sharing opportunity, however, but instead hold large shares of their portfolios in domestic equities. Earlier studies in the literature mainly provide theoretical explanations for the home bias. One major explanation is that there are barriers to international investment such as governmental restrictions on foreign and domestic capital flows, foreign taxes, and high transaction costs associated with international investments (Errunza and Losq 1985). Many financial economists, however, have found that even though the barriers to international investment have fallen dramatically, foreign owner-

ship of shares is still extremely limited and much smaller than one would expect in the absence of these barriers. Home bias is prevalent even among investors with relatively low costs of transacting in financial markets such as institutional investors. Other explanations include investors' desire to hedge against inflation risk, deviations from purchasing power parity (Adler and Dumas 1983; Cooper and Kaplanis 1994), and the price uncertainty of human capital or other nontraded assets (Eldor, Pines, and Schwartz 1988; Stockman and Dellas 1989). By holding sufficient amounts of domestic assets, risk-averse investors may hedge against price uncertainty of the goods they consume, since the average return on the portfolio is high (low) when consumption becomes expensive (inexpensive). None of these factors, however, can fully explain the observed reluctance of investors to diversify their holdings internationally.

An alternative view attributes home bias to empirical mismeasurement of the home bias. Some researchers find that, after incorporating empirical uncertainty into the analysis, there may be no home bias because foreign diversification does not lead to a statistically significant improvement in portfolio performance, though the general validity of this thesis remains in doubt (Bekaert and Urias 1996; Pastor 2000).

As investor behavioral factors have been proposed to explain various anomalies in finance since the 1990s, familiarity has become the most dominant explanation of the home country bias puzzle. Familiarity is associated with a general sense of comfort with the known and discomfort with, or even fear of, the alien and distant, which adds a nonpecuniary dimension to the traditional risk-return trade-off emphasized by portfolio theory. Huberman (2001) finds that familiarity breeds investment, and paucity of international diversification is only one of the implications of this tendency to invest in the familiar. Familiarity can feed into the home bias puzzle through two channels: "pure" familiarity effect due to geographic proximity (such as language, culture, and distance) and, more important, information asymmetries caused by distance, which include low

visibility of foreign firms, an inability to monitor the firms, and low credibility of financial information (Kang and Stulz 1997). Separating out pure familiarity effect from informational asymmetries, however, remains a challenge for the empirical literature. Some studies document that even within their foreign portfolios   however small these foreign holdings may be   investors prefer the stocks of foreign countries that are closer (Portes and Rey 2005) and whose equity markets are more, not less, correlated with their own (Chan, Covrig, and Ng 2005). Other studies find that familiarity-caused home bias is not only international, but also regional. For instance, U.S. investment managers exhibit a strong preference for locally headquartered firms in their domestic portfolios, where local investments reflect a true informational advantage (Coval and Moskowitz 1999). In addition, in Finland, there is evidence that investors are more likely to hold and trade stocks of Finnish firms that are located close to them, that communicate in their native language, and that have chief executives of the same cultural background (Grinblatt and Keloharju 2001).

Although the various factors just mentioned have been shown in the literature to explain the home bias to some extent, no single explanation seems to fully account for the extent of home bias. Nevertheless, attempts to explain home bias have helped raise questions to guide further studies on the issue.

Home Country Bias in Macroeconomics The home country bias also applies to real, not just financial, investments. Indeed, the oft-documented high positive correlation between a country's long-term savings rate and its investment rate suggests that money saved in a country is invested in that country rather than in the (possibly foreign) country that offers the best return. A closely related phenomenon is the so-called consumption home bias in macroeconomics, which refers to the lack of risk sharing observed in consumption co-movements across countries. Intuitively, the equity home bias puzzle and consumption home bias puzzle may be linked, yet home bias in equities is neither sufficient nor necessary for home bias in consumption (Lewis 1999). On the other hand, home bias in portfolio holdings and home bias in consumption may have some common causes such as various costs of international transactions (Obstfeld and Rogoff 2001). Further, understanding about the potential link between the two puzzles may rely on a better unified understanding of consumption and equity prices.

See also asymmetric information; balance of payments; capital flows to developing countries; capital mobility; Feldstein-Horioka puzzle; hedging; purchasing power parity

FURTHER READING

Adler, M., and B. Dumas. 1983. "International Portfolio Choice and Corporate Finance: A Synthesis." *Journal of Finance* 38 (3): 925 84.

Bekaert, G., and M. S. Urias. 1996. "Diversification, In tegration, and Emerging Market Closed end Funds." *Journal of Finance* 51 (3): 835 69.

Chan, K., V. Covrig, and L. Ng. 2005. "What Determines the Domestic Bias and Foreign Bias? Evidence from Mutual Fund Equity Allocations Worldwide." *Journal of Finance* 60 (3): 1495 1534.

Cooper, I., and E. Kaplanis. 1994. "Home Bias in Equity Portfolios, Inflation Hedging, and International Capital Market Equilibrium." *Review of Financial Studies* 7 (1): 45 60.

Coval, J., and T. Moskowitz. 1999. "Home Bias at Home: Local Equity Preference in Domestic Portfolios." *Journal of Finance* 54 (6): 2045 73.

Eldor, R., D. Pines, and A. Schwartz. 1988. "Home Asset Preference and Productivity Shocks." *Journal of International Economics* 25 (1 2), 165 76.

Errunza, V., and E. Losq. 1985. "International Asset Pricing and Mild Segmentation: Theory and Test." *Journal of Finance* 40 (1): 105 24.

French, K., and J. Poterba. 1991. "Investor Diversification and International Equity Markets." *American Economic Review* 81 (2): 222 26.

Grinblatt, M., and M. Keloharju. 2001. "How Distance, Language, and Culture Influence Stockholdings and Trades." *Journal of Finance* 56 (3): 1053 73.

Huberman, G. 2001. "Familiarity Breeds Investment." *Review of Financial Studies* 14 (3): 659 80.

Kang, J. K., and R. M. Stulz. 1997. "Why Is There a Home Bias? An Analysis of Foreign Portfolio Equity Owner

ship in Japan." *Journal of Financial Economics* 46 (1): 3 28.

Levy, H., and M. Sarnat. 1970. "International Diversifica tion of Investment Portfolios." *American Economic Re view* 60 (4):668 75.

Lewis, K. K. 1999. "Trying to Explain Home Bias in Equities and Consumption." *Journal of Economic Lit erature* 37 (2): 571 608.

Obstfeld, M., and K. Rogoff. 2001. "The Six Major Puzzles of International Macroeconomics Solved." In *National Bureau of Economic Research Macroeconomics Annual*, edited by B. Bernanke and K. Rogoff. Cambridge, MA: MIT Press, 339 90.

Pastor, L. 2000. "Portfolio Selection and Asset Pricing Models." *Journal of Finance* 55 (1): 179 223.

Portes, R., and H. Rey. 2005. "The Determinants of Cross border Equity Flows." *Journal of International Economics* 65 (2): 325 51.

Stockman, A. C., and H. Dellas. 1989. "International Portfolio Nondiversification and Exchange Rate Varia bility." *Journal of International Economics* 26: 271 89.

Tesar, L., and I. Werner. 1998. "The Internationalization of Securities Markets since the 1987 Crash." In *Brookings Wharton Papers on Financial Services*, edited by R. Litan and A. Santomero. Washington, DC: The Brookings Institution, 281 372.

FANG CAI

■ **host country**

See location theory

■ **hot money and sudden stops**

Hot money, also known as *mobile capital*, refers to short-term private capital that flows across borders into different markets in search of higher returns and can be easily reversed. Hot money consequently makes a country susceptible to "sudden stops." The sudden stop problem, first emphasized by Calvo (1998), refers to an abrupt cessation in foreign capital inflows and/or a sharp capital outflow concurrent with a currency/balance-of-payment crisis. Notable episodes of sudden stop crises are the financial crises that hit Mexico in 1994, Thailand and Korea in 1997, Indonesia, Malaysia, and Russia in 1998, and Argentina in 2001.

There is a growing body of literature that attempts to define and measure sudden stops. One definition of a sudden stop is any episode in which two con ditions are met: (1) there is a significant reversal of capital inflows (a decrease in the financial account of at least 2 standard deviations below the country-specific sample mean), and (2) the current account deficit is reduced in either year t or in $t+1$ (Calvo, Izquierdo, and Talvi 2006). Another definition of a sudden stop is an event in which a currency crisis and a current account reversal (an increase in the current account surplus of more than 3 percent of GDP) occur simultaneously (Hutchison and Noy 2006).

What leads to a sudden stop? Even after ac counting for all macroeconomic fundamentals that may cause a capital outflow, a large unexplained component remains, which has been linked to herding behavior (Chari and Kehoe 2003).

Sudden Stops and Output Contractions Re gardless of the exact definition, sudden stops lead to a large, though short-lived, negative impact on output growth that is substantially greater than the impact associated with a currency or financial crisis without a sudden stop. This abrupt reversal in foreign credit inflows, in conjunction with a realignment of the exchange rate, typically causes a sharp drop in do mestic investment, domestic production, and em ployment. The recovery from these types of crises seems to be rapid, though it is usually not accom panied by full resurgence in capital inflows or by a recovery of domestic investment. There is evidence that sudden stops generally have a larger adverse impact on output in countries relatively closed to trade and with greater exchange rate fixity (Edwards 2004).

The question that follows is: What are the exact mechanics by which mobile capital leads to a finan cial crisis? The crisis-inducing nature of bank loans/ debts is based on an open economy version of the bank panic model (Diamond and Dybvig 1983). Following some negative shock, depositors, con cerned about the safety of their savings, attempt to

withdraw en masse (which occurs because, given the "first-come-first-served" rule of deposit withdrawals, depositors all rush to withdraw their money before the banks fail). Since the banks' liquid asset/reserves (cash available for withdrawals) are often less than their potential obligations (demand deposits), they are forced into the premature liquidation of long-term investments. Given the partial irreversibility of investments, they obtain a lower return on liquidation. In the open economy, if the foreign currency revenues obtainable in the short term are less than the corresponding short-term potential foreign currency obligations, the banks are "internationally illiquid" (Chang and Velasco 1999), while creditors, such as foreign banks, are unwilling to roll over short-term interbank loans, especially for emerging markets. The sudden termination of bank finance that follows forces the abandonment of potentially solvent investment projects. This consequent decline in capital formation indeed, capital destruction leads to a sudden output/economic collapse.

The maturity mismatch story leading to a possible bad equilibrium in the event of a bank panic is well known, but there are many other possible channels of output contraction as well. Another prominent explanation relies on a "Fisherian" debt-deflation channel (Fisher 1933). In these models, deflation, triggered by an endogenous occurrence of a sudden stop when credit constraints become binding in high-debt states, leads to lower marginal returns to factors of production; this leads, through a debt-deflation amplification process, to a large decline in economic activity. The debt-deflation multiplier can be even more severe: firm bankruptcies may cause banks to become ever more cautious, reducing lending to other firms, and thus inducing a further fall in credit (Mendoza 2006).

Adverse balance sheet effects an increase in liabilities that is not matched by a corresponding increase in assets, due to a combination of currency depreciation following a reversal in hot money and unhedged (uninsured) borrowing in foreign currency are another channel through which sudden stops could adversely affect the real economy (Rajan and Shen 2006).

Importance of Financing via Foreign Direct Investment Economies that finance their current account deficits mainly through foreign direct investment (FDI) generally are less susceptible to sudden stops than are economies that experience high levels of hot money inflows. Thus, Hausmann and Fernández-Arias (2000) refer to short-term debt as "bad cholesterol" as it is motivated by "speculative considerations" such as exchange rate expectations. This type of financing is the first to exit in times of trouble. The resulting boom-bust cycle of capital flows in the 1990s inflicted great damage on many emerging economies.

Of course, short-term debt is not the only form of liquid liability. An alternative and more complete measure of illiquidity is mobile capital, or international capital markets, which refers to short-term bank loans plus portfolio investment in the form of equity and bond issues in offshore markets. Equity investment at least shares in the risk of falling market values, while long-term bond issues are less liquid. But FDI is seen as "good cholesterol" because it is bolted down and is relatively irreversible in the short run. It flows in because it is attracted by long-term prospects; it enhances the productive capacity of the country and produces the revenue streams necessary to cover future capital outflows (if they occur) without increasing the overall indebtedness of the economy.

Available summary statistics of private capital flows to developing economies suggest, consistent with conventional wisdom, that FDI has been the most resilient form of external financing (for instance, see Chuhan et al. 1996; Sarno and Taylor 1999). Empirical analysis suggests that emerging economies most prone to currency crashes tend to have a relatively smaller share of FDI in total capital inflows and a relatively higher share of short-term external debt (Frankel and Rose 1996). Other studies have confirmed that short-term indebtedness is a robust predictor of financial crises (Rodrik and Velasco 1999).

A potential criticism of the conventional view regarding differing degrees of stability of various capital flows is that complex interactions between

FDI and other flows are not taken into account. Therefore examining each flow individually, particularly during short periods of time (such as year-to-year variations) may not be a reliable indicator of the degree of risk of various classes of flows at best and could be highly misleading at worst (see Bird and Rajan 2002, and references cited there).

Contrary to popular belief, FDI itself, which is a form of external finance, is not "bolted down," though the physical assets it finances are. Foreign investors could use the physical assets as collateral to obtain a loan from banks and place the funds abroad. In other words, the foreign direct investor may hedge the firm's FDI exposure by borrowing domestically and taking short-term capital out of the country. Hence a firm may be doing one thing with its assets and a completely different thing with the manner in which it finances them. This appears consistent with the Malaysian capital flows data during the crisis where, as noted, portfolio outflows in 1997 sharply outweighed the cumulative inflows between 1980 and 1996. Apparently the portfolio outflows must have entered from some other account (such as FDI or bank loans) (Bird and Rajan 2002). Indeed, the distinction between portfolio and FDI flows in the balance of payments can be somewhat arbitrary, and the proportion of FDI flows in aggregate capital flows may be overstated. Small differences in equity ownership are unlikely to represent substantially different investment horizons.

All of this suggests that the casual presumption that the switch from hot money to FDI alone will automatically safeguard a country against sudden stops and output contractions should be viewed with a degree of caution.

See also balance of payments; balance sheet approach/effects; bubbles; capital controls; capital flows to developing countries; currency crisis; debt deflation; financial crisis; foreign direct investment (FDI); hedge funds; interest parity conditions; speculation; spillovers

FURTHER READING

Bird, Graham, and Ramkishen S. Rajan. 2002. "Does FDI Guarantee the Stability of International Capital Flows? Evidence from Malaysia." *Development Policy Review* 20 (2): 191 202. Examines the links between FDI and currency crisis with particular reference to Malaysia in 1997 98.

Caballero, Ricardo, and Arvind Krishnamurthy. 2004. "Smoothing Sudden Stops." *Journal of Economic Theory* 119 (1): 104 27. Describes various financial polices that can potentially prevent the occurrence of sudden stops.

Calvo, Guillermo. 1998. "Capital Flows and Capital Market Crises: The Simple Economics of Sudden Stops." *Journal of Applied Economics* 1 (1): 35 54. The seminal work in the economic literature that describes sudden stop crises from a theoretical perspective.

Calvo, Guillermo A., Alejandro Izquierdo, and Ernesto Talvi. 2006. "Sudden Stops and Phoenix Miracles in Emerging Markets." *American Economic Review* 96 (2): 405 10. An empirical examination of the regularities of sudden stops, according to which gross domestic product typically recovers quickly following the crisis; this recovery, however, is not accompanied by resurgence of capital flows.

Chang, R., and A. Velasco. 1999. "Liquidity Crises in Emerging Markets: Theory and Policy." In *NBER Macroeconomics Annual 1999*, edited by B. Bernanke and J. Rotemberg. Cambridge, MA: MIT Press, 11 78. Develops a simple model to explain how short term debt can lead to currency crisis.

Chari, V. V., and Patrick J. Kehoe. 2003. "Hot Money." *Journal of Political Economy* 111 (6): 1262 92. Develops a model of herd behavior to explain volatile capital flows.

Chuhan, P., G. Perez Quiros, and H. Popper. 1996. "International Capital Flows: Do Short term Investment and Direct Investment Differ?" Policy Research Working Paper No. 1507. Washington, DC: World Bank. Examines whether there are differences in FDI and short term capital flows.

Diamond, P., and P. Dybvig. 1983. "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy* 91 (3): 401 19. Develops a bank based model on liquidity crises.

Edison, Hali, and Carmen Reinhart. 2001. "Stopping Hot Money: On the Use of Capital Controls during Financial Crises." *Journal of Development Economics* 66: 533 53. Evaluates the effectiveness of capital controls in

preventing speculative reversal of hot money in emerging economies.

Edwards, Sebastian. 2004. "Thirty Years of Current Account Imbalances, Current Account Reversals, and Sudden Stops." NBER Working Paper No. 10276. Cambridge, MA: National Bureau of Economic Research. Examines the empirical evidence relating to sudden stops.

Fisher, Irving. 1933. "The Debt Deflation Theory of Great Depressions." *Econometrica* 1 (4): 337 57.

Frankel, J., and A. Rose. 1996. "Currency Crashes in Emerging Markets: Empirical Indicators." *Journal of International Economics* 41 (3 4): 351 68.

Hausmann, R., and E. Fernández Arias. 2000. "Is FDI a Safer Form of Financing?" Working Paper No.416. Inter American Development Bank. Critically examines whether FDI is in fact a stable and safe source of financing.

Hutchison, Michael, and Ilan Noy. 2006. "Sudden Stops and the Mexican Wave: Currency Crises, Capital Flow Reversals, and Output Loss in Emerging Markets." *Journal of Development Economics* 79 (1): 225 48. Empirically defines what sudden stops are and measures the depth of the recessions that typically follow these crises.

Mendoza, Enrique G. 2006. "Lessons from the Debt Deflation Theory of Sudden Stops." *American Economic Review* 96 (2): 411 16. Describes a set of dynamic, stochastic general equilibrium economic models that replicate many of the defining characteristics of sudden stop crises.

Rajan, Ramkishen S., and Shen Chung Hua. 2006. "Why Are Crisis Induced Devaluations Contractionary? Exploring Alternative Hypotheses." *Journal of Economic Integration* 18 (1): 1 24. Empirical investigation of the various factors that might lead a currency depreciation to be contractionary rather than expansionary.

Rodrik, D., and A. Velasco. 1999. "Short term Capital Flows." NBER Working Paper No. 7364. Cambridge, MA: National Bureau of Economic Research. Examines the impact of short term capital flows on currency crisis.

Sarno, L., and M. Taylor. 1999. "Hot Money, Accounting Labels, and the Permanence of Capital Flows to Developing Countries: An Empirical Investigation." *Journal of Development Economics* 59 (28): 337 64. Examines the permanence/durability of various types of capital inflows.

Williamson, John. 2005. *Curbing the Boom Bust Cycle: Stabilizing Capital Flows to Emerging Markets.* Washington, D.C.: Institute for International Economics. This policy brief addresses many of the possible policy solutions to the sudden stop problem that has plagued financial markets in emerging markets for the past two decades.

**ASHIMA GOYAL, ILAN NOY,
AND RAMKISHEN S. RAJAN**

■ illegal drugs trade

Despite the lack of reliable estimates on other illegal markets, the trade in prohibited narcotics and other psychoactive drugs is generally considered to be the illegal activity with the largest turnover worldwide. The *2005 World Drug Report* of the United Nations Office on Drugs and Crime (UNODC) includes a systematic effort to produce a comprehensive set of estimates. For illicit drugs as a whole, the UNODC estimates a total of almost $322 billion in retail sales in 2003; $94 billion in wholesale revenues; and $13 billion in producer sales (UNODC 2005).

The largest market, according to the UN calculations, is cannabis herb (normally called marijuana), which has a retail market size of $113 billion. Marijuana is followed by cocaine ($71 billion), the opiates ($65 billion), and cannabis resin (popularly called hashish, with $29 billion). The markets for synthetic drugs (properly called amphetamine-type stimulants and including methamphetamine, amphetamine, and ecstasy) total about $44 billion. Though derived from an economic model, these figures are in reality guesstimates and have to be treated with great caution. The basis inputs that are needed for such calculations data on production, prices, and quantities exported, imported, and consumed are themselves often estimates and frequently based on deficient data.

If compared to global licit exports or global gross domestic product (GDP) (respectively, $7,503 billion and $35,765 billion in 2003), the estimated size of the global illicit drug market may not appear to be very large. Total retail drug expenditures correspond to 0.9 percent of global GDP and drug wholesale revenues represent only 1.3 percent of global exports measures. The size of the global illicit drug market is not insubstantial, however. Illicit drug wholesale revenues account for 14 percent of global agricultural products and are much higher than the export value of most licit agricultural commodities (UNODC 2005).

The global drug market also involves a considerable number of people. According to the 2006 UNODC estimates, about 200 million people currently use illicit drugs at least once a year (about 5 percent of the world population age 15 64). Cannabis is by far the most popular and widespread illegal drug: out of the 200 million total, 162 million are cannabis users. Cannabis is followed by synthetic drugs (some 35 million people), which include amphetamines (used by 35 million people) and ecstasy (almost 10 million people). The number of opiate users is estimated at some 16 million people, of which 11 million are heroin users. Some 13 million people use cocaine at least once a year. A large proportion of drug users can be found in Asia. More than 60 percent of the world's amphetamine users and more than half of opiate users reside there (UNODC 2006; Paoli, Greenfield, and Reuter 2008). Asia accounts only for a small portion of total retail expenditures, however: $35 billion or 11 percent. With $142 billion (or 44 percent) and $106 billion (or 33 percent), respectively, North America and Europe account for the bulk of total retail expenditure

estimates (UNODC 2005), simply because retail drug prices are much higher there than in developing nations.

Drugs and Source Countries The two most dangerous plant-based drugs, heroin and cocaine, are produced in a small number of poor countries, which are often thousands of miles away from many end-users. Virtually all the cultivation of the coca bush (from which cocaine is made) is concentrated in three Andean countries: Colombia, Peru, and Bolivia, accounting for 63, 25, and 11 percent respectively of world production in 2005. Most cocaine hydrochloride is processed in Colombia and from there is exported into the United States and Europe (UNODC 2006).

From the late 1980s on, Afghanistan and Burma effectively "owned" world opium production, with a fringe of second- and third-tier producers contributing only modestly. Together, Afghanistan and Burma (or Myanmar, as it was renamed by the military dictatorship in 1989) accounted for more than 95 percent of world production in 2005. After 2001, however, the two-country dominance began shifting to one-country dominance, with Afghanistan accounting for an increasing share of world production Afghanistan's share of the total was more than 88 percent in 2005. Although it accounted in 2005 for less than 1 percent of world's heroin production, Colombia has since the mid-1990s become one of the major suppliers of the U.S. market. With slightly higher figures (1.5 percent of total production), Mexico is also an important source of heroin for the U.S. market (UNODC 2006; Paoli, Greenfield, and Reuter 2008).

The cultivation of cannabis is much more decentralized than that of opium and coca so much so that in 2006 the UNODC identified 176 countries or territories where cannabis is produced. This spread is primarily due to the great adaptability of the cannabis plant, which grows well in virtually every inhabited region of the world and can be cultivated with little maintenance in small plots or even indoors. In addition, unlike most other illicit drugs, cannabis products can be consumed with little processing after harvesting. Thanks to these two char-

acteristics, users can feasibly cultivate their own supply even in countries that rigorously apply the world prohibition regime. The U.S. authorities, for example, report that about two-thirds of the cannabis consumed in the country, predominantly in the form of marijuana, is domestically produced. The main foreign supplier of marijuana in the United States is Mexico. Hashish, the cannabis resin that is particularly popular in Europe, predominantly comes from Morocco (UNODC 2006).

Even more than cannabis, synthetic drugs are rarely trafficked across long distances because they can be easily manufactured in small and mobile laboratories in, or close to, areas of consumption. Hence, for example, the majority of the ecstasy pills consumed in Europe is produced in the Netherlands and, to a much lesser extent, in Belgium and some Eastern European countries. Likewise, the U.S. demand for methamphetamine is traditionally supplied by clandestine laboratories in California, several Midwestern states, and Mexico. Reflecting Asia's lion's share of the synthetic drug market, Burma is the world largest producer of methamphetamine (UNODC 2006).

Evolution of Drug Markets and Variable Characteristics of Suppliers Notwithstanding the large number of illegal drug users worldwide and the magnitude of the world drug economy, the illegal drug trade is a relatively new phenomenon. In fact, only in the last three decades of the 20th century did it assume mass proportions. Until about the beginning of the 20th century, all psychoactive drugs could be freely produced and consumed not only in Asia and Latin America, but also in Western countries. The production, trade, and consumption of opium, coca, and their derivatives began to be severely regulated and then largely prohibited (except for limited medical purposes) only after the first International Opium Convention was concluded in The Hague in 1912. The new international drug control regime, whose enforcement was entrusted after World War I to the League of Nations and after World War II to the United Nations, seemed quite successful at first. During the 1920s, 1930s, and 1940s the consumption of all prohibited substances rapidly declined in

the United States, Europe, and many Asian nations (scholars debate the extent to which the interwar decline is attributable to increasing prohibition rather than other factors, above all, changing medical and public attitudes toward drugs; see Courtwright 2001; Berridge 1999). Under pressure from the United States, cannabis trade and possession were also increasingly restricted and, finally, subjected to the international drug control regime in the early 1960s.

From the 1960s onward, the demand for heroin, cocaine, and cannabis rose again, first in the United States and then in Western Europe. At the end of the 20th century, it also grew in the second (former communist) and third-world countries. The postwar expansion of illicit drug use was caused by contingent events such as the Vietnam War, which brought thousands of young American soldiers into contact with heroin and macrosocial changes. Among the latter, two are most important: (1) the rise of a youth mass subculture, which resorted to illegal psychoactive substances to distinguish itself from the mainstream culture, and (2) technological progress, which made communication, travel, and trade in both legal and illegal commodities easier and faster.

From the 1960s onward, the rising demand for illicit drugs fostered the development of drug distribution systems able to transfer drugs from producers to consumers. In the beginning, illegal drugs were imported by the consumers themselves, who used some of them and sold the rest within a close circle of friends. It was a sort of "ants trafficking." Soon, however, in both the United States and Western Europe the professional role of the drug dealer began to consolidate. In a few years, the development of a large-scale drug market fostered the progressive entry of professional criminals into the drug business.

In source countries, where the state authorities are often unable or unwilling to enforce the international prohibition regime, large organizations have sometimes emerged to coordinate opium poppy and coca cultivation and to process heroin and cocaine. In northern Burma, for example, the refinement and export of heroin were first organized by the nationalist Chinese Kuomintang Army and, more recently,

by several armies representing local ethnic minorities. Much of the coca cultivation in Colombia also takes place in the inland areas controlled by guerrilla and paramilitary movements, particularly the FARC (Fuerzas Armadas Revolucionarias de Colombia).

In developed countries, the constraints deriving from the illegal status of the drugs have so far prevented the consolidation of large-scale, hierarchically organized drug-trafficking enterprises. These constraints arise from two facts: all illegal market actors particularly drug traffickers and dealers are obliged to operate (1) without the state and (2) against the state.

Since the goods and services they provide are prohibited, illegal market suppliers operate without the state; they cannot resort to state institutions to enforce contracts and have the violations of contracts prosecuted. As a result, property rights are poorly protected, employment contracts can hardly be formalized, and the development of large, formally organized, enduring companies is strongly discouraged.

All suppliers of illegal commodities specifically drugs are also required to operate against the state, that is, under the constant threat of being arrested and having their assets confiscated by law enforcement institutions. They therefore try to organize their activities to assure that the risk of police detection is minimized. Incorporating drug transactions into kinship and friendship networks and reducing the number of customers and employees are two of the strategies that drug entrepreneurs most often employ to reduce their vulnerability to law enforcement efforts (Reuter 1985).

Empirical research confirms these insights. In Europe and North America, the great majority of drug deals, even those involving large quantities, are carried out by numerous, relatively small, and often ephemeral enterprises. Especially at the intermediate and lower levels, many dealers work alone, either to finance their own drug habits or, more rarely, to earn fast money. Many drug distributors are members of ethnic minorities who either exploit direct connections to source countries or are pushed into dealing by lack of rewarding alternatives in the legitimate

economy. At the retail level and in closed settings, drugs are also often distributed by inconspicuous persons who have no contact with the underworld.

For the same reasons, the relationships among illegal drug enterprises generally involve competition rather than collusion. The best evidence against control is simply the ease with which new sellers enter and the speed with which dealers depart. There may be rents for various capacities, but certainly no power to exclude. Throughout Europe and North America, moreover, drug-dealing firms are price-takers rather than price-givers: that is, none of them is able to influence the commodity's price appreciably by varying the quantity of the output it sells. The continuing decline of prices during the 20-year period 1985 2005 at all levels of the market in heroin and most other drugs suggests that, if markets power ever existed, it has by now been dissipated.

The danger that the contemporary illegal drugs industry represents for the world economy is not limited to its turnover and the number of people it involves as producers, traffickers, and users. Two further aspects need to be considered. The industry is a major source of corruption, violence, and instability in a number of drug-producing and transit countries its collateral effects being occasionally reinforced by the war on drugs itself. Moreover, the international routes and networks that have been developed for the transportation of illegal drugs from source countries to the final consumer nations (and for bringing back the money the other way round) can be used for a plurality of other illicit goals, ranging from human smuggling and trafficking to terrorism financing.

See also globalization; primary products trade; smuggling

FURTHER READING

Berridge, Virginia. 1999. *Opium and the People: Opiate Use and Drug Control Policy in Nineteenth and Early Twentieth Century England*. London: Free Association Books. The only empirical account of the history of opiate use and drug control policy in a European nation.

Courtwright, David T. 2001. *Dark Paradise: A History of Opiate Addiction in America*. Cambridge: Harvard University Press. The standard work on the history of opiate use in the United States from the mid nineteenth century to the 1990s.

Paoli, Letizia, Victoria Greenfield, and Peter Reuter. 2008. *The World Heroin Market: Can Its Supply Be Reduced?* New York: Oxford University Press. A seminal analysis of the evolution, composition, and behavior of the world opiate market and an objective assessment of what nations and the international community can achieve in drug control.

Reuter, Peter. 1985. *The Organization of Illegal Markets: An Economic Analysis*. Washington, DC: National Institute of Justice. A compact but very insightful analysis of the organization of illegal markets in developed countries.

Thoumi, Francisco E. 2003. *Illegal Drugs, Economy, and Society in the Andes*. Washington, DC: Woodrow Wilson Center. A seminal analysis of the cocaine industry in three main producing countries Colombia, Bolivia, and Peru and its impact upon the local legitimate economy and society.

UNODC (United Nations Office on Drugs and Crime). Annual. *World Drug Report*. Vienna: UNODC. Downloadable from http://www.unodc.org/unodc/en/world drug report.html. The most comprehensive annual description of the world illicit drug markets.

LETIZIA PAOLI

■ illiquidity

See banking crisis

■ import substitution industrialization

Import substitution industrialization (ISI) is a trade and economic policy based on the premise that a developing country should attempt to substitute products it imports with domestically produced substitutes. The policy has three major tenets: (1) an active industrial policy to promote a domestic industrial base producing strategic substitutes, often involving strong public sector investment in infrastructure and in "strategic" sectors, and the establishment of development banks to support these ac-

tivities; (2) protective barriers to trade (namely, tariffs and quotas to protect new or infant industries) and changes in the internal terms of trade against traditional primary exports; and (3) a monetary policy that rations foreign exchange, with multiple exchange rates to channel it preferentially to the imports of noncompetitive intermediate and capital goods. Generally, a distinction is made between a first "easy" phase of ISI, during which mostly nondurable goods are produced, and later more "mature" phases in which import substitution deepens, producing nondurable consumer as well as intermediate and capital goods, and (in some cases) more emphasis is given to export promotion (by the imposition of export incentives on old layers of import protection).

Historical Origins The first experiments with ISI were launched and sustained by the industrializing countries of Europe, North America, and Russia in the 19th century in order to overcome their relative economic (and political) backwardness vis-à-vis the industrial leader, then the United Kingdom (Gerschenkron 1962; Chang 2002). Early theoretical justifications for ISI include List (1841) and Manoïlesco (1928). The most coherent post World War II formulations of and justifications for ISI appear in Prebisch (1950), Nurkse (1959), and Gerschenkron (1962). Important reviews of the evolution of ideas about ISI are Bruton (1998) and Waterbury (1999).

The Rationale for ISI Import substitution policies were adopted by many low-income developing nations during the early 1950s as a means to escape from the primary commodity specialization trap and to promote industrial diversification. Specialization in labor-intensive primary commodities was considered undesirable because of declining world markets for primary products, the limited scope for technological progress in primary production, and the long-term deteriorating terms of trade of primary commodities, which caused low-income countries to have to pay relatively more for manufactured goods imported from industrialized countries. Further motivation for ISI was provided by the lack of adequate external financing, in combination with a reluctance to devalue the exchange rate (for fear of its inflationary effects), and the high levels of (agricultural) protection in the industrialized countries.

Industrialization of the developing world was considered imperative: only by the transformation of productive structures toward manufacturing could average labor productivity be raised so as to generate a higher standard of living (through increased real wages) and to reduce the incidence of poverty and destitution. Crucial to the generation of increased (manufacturing) productivity growth was the development of substantial technological capabilities, including technical knowledge, quality control practices, entrepreneurial and managerial skills, and work habits (Amsden 2001). The buildup of such technological capabilities is a path-dependent process in which accumulated experience plays the crucial role; it involves learning by doing, adaptation, and induced technological change as well as the development of manufacturing entrepreneurial and working classes. ISI was regarded as essential in kick-starting the development of such technological capabilities. While the initial costs of industrial production are higher than former import prices, the economic rationale for ISI is that, given time, (1) the industry will eventually be able to reap the benefits of large-scale production, technological learning, manufacturing experience (learning by doing), and lower production costs (the so-called infant industry argument for tariff protection), and (2) the balance of payments will be improved as fewer manufactured goods are imported.

ISI Policy Instruments The typical strategy is first to erect tariff barriers or quotas on the importation of strategic commodities and then to try to set up a domestic industry to produce the goods imported; often, this involves cooperation with foreign companies encouraged to set up their plants behind the wall of tariff protection and given various tax and investment incentives. Tariff protection is supplemented by numerous quantitative and direct controls on the availability of foreign exchange. The other major policy instrument is the exchange rate, which is generally overvalued (relative to a free trade situation) in order to keep the domestic price of

capital goods low. Often, *multiple* exchange rate systems are used instead of trade policy to implicitly tax competitive imports and traditional exports and subsidize complementary imports. Domestically, ISI policies include strong public-sector investment in infrastructure and in strategic sectors, including energy and irrigation, the establishment of development banks to support these activities, industrial policy regulating domestic investment, and price and fiscal policies to change the domestic terms of trade against agriculture and traditional primary exports (Bruton 1998; Amsden 2001).

The Logic behind ISI In theory, there is no presumption that ISI will unambiguously reduce or accelerate economic growth in relation to openness in trade. The main complications are twofold.

First, in traditional (static) trade theory, assuming that no market imperfections and other imposed distortions exist, the effect of import protection (as under ISI) is to reduce the level of real gross domestic product (GDP) at world prices; however, simple Harberger triangles from a competitive model rarely identify welfare losses from trade restrictions larger than 1.5–2 percent of GDP, which is far too small to produce significant changes in growth (Ocampo and Taylor 1998; Winters 2004). But in the presence of market failures (such as positive production externalities in import-competing sectors), traditional trade theory shows that real GDP can be higher under protection than under free trade (Bhagwati 1958).

Second, in models of endogenous growth, generated by nondiminishing returns to reproducible factors of production or by learning by doing and other variants of technological progress, the forces of comparative advantage may push an economy's resources in the direction of dynamic activities that generate long-run growth (via externalities in research and development, expanding product variety, learning by doing, increasing returns to scale, agglomeration effects, the creation of forward and backward linkages, and so on) *or divert* them from such activities into traditional, nondynamic ones. Free trade can thus be detrimental to a country's economic growth prospects, especially when it lacks

adequate technological capabilities and (given its initial factor endowments) has a comparative advantage in nondynamic sectors. Import protection may lead to higher long-run growth of real GDP if the trade restrictions promote technologically more dynamic sectors over others (Ocampo and Taylor 1998; Rodríguez and Rodrik 2001).

Experience with ISI The voluminous empirical literature evaluating the benefits and costs of ISI can be classified into two categories: cross-country regression-based studies investigating the impact of import protection (often under ISI) on economic growth (Rodríguez and Rodrik 2001; Srinivasan and Bhagwati 2001; Winters 2004), and in-depth (comparative) analyses of historical country experiences (Little et al. 1970; Balassa et al. 1971; Bhagwati 1978; Krueger 1978; Cardénas et al. 2000; Amsden 2001; Wade 2004). From these two strands of literature, no general, unambiguous conclusion about the *net* impact of ISI policies on growth (and economic performance in general) emerges, but only *contingent* conclusions pointing to a large number of country-specific, often political-economy, and external characteristics.

Cross-Country Regression Studies Surveys of cross-national econometric studies of the relationship between trade policy regimes (including ISI trade policy regimes) and growth, including Edwards (1993), Rodríguez and Rodrik (2001), and Winters (2004), conclude that much of the cross-country (linear) regression studies have been plagued by conceptual and empirical shortcomings, lack of relevant control variables, problems of causality and endogeneity, multicollinearity, lack of appropriate control variables, and measurement errors. As a result, no robust systematic conclusions about the impact on growth of ISI trade policies can be drawn from this literature (Srinivasan and Bhagwati 2001).

Comparative Country Studies The effects of ISI have been analyzed in a number of influential studies under the auspices of the Organisation for Economic Co-operation and Development, the National Bureau of Economic Research, and the World Bank (e.g., Little et al. 1970; Balassa et al. 1971; Bhagwati 1978; Krueger 1978). These identified substantial

macroeconomic costs of ISI. First, ISI led to inefficiencies in resource allocation, because through an overvalued exchange rate it introduced a bias against exports and in favor of (capital-intensive sectors in) the home market, thus leading to underutilization of capital stock (in capital-scarce economies), declining capital productivity, and a failure of high investment rates to significantly reduce widespread unemployment and underemployment (Bruton 1998). In addition, ISI generally gave rise to a wide and nontransparent dispersion of effective exchange rates, which served to make the protectionist regime more complex, distorting the structure of incentives that ISI itself required. Second, imports generally rose faster than expected due to the demand for capital goods and intermediate goods to support the new industries; as a result, balance of payments problems were widespread the more so because (agricultural) exports were penalized by the broad strategy of ISI (Little et al. 1970; Balassa et al. 1971). Rather than reducing dependence on imported inputs (energy) and technology, the strategy may have significantly increased it. Third, IS regimes have been argued to induce directly unproductive profit-seeking (DUP) activities (Bhagwati 1982), diverting resources from productive use into unproductive but profitable lobbying to change or evade policies or to seek the revenues they generate; this, in turn, reduced investment and productivity growth and hence long-term growth (Winters 2004). Fourth, in countries where the home market is relatively small, ISI created less competitive markets, which in some cases reduced aggregate efficiency, productivity growth, and innovativeness. Taken together, these tendencies suggest that the trade and exchange rate policies of ISI became self-defeating in terms of its objectives: to increase capital accumulation and generate a more diversified productive structure of production.

During the 1990s, however, the impact of ISI on economic performance was reconsidered (Bruton 1998; Rodrik 1999) particularly within the context of the rapidly growing East Asian economies (Storm and Naastepad 2005). According to this "revisionist" literature, ISI worked well during the period 1950 73, bringing unprecedented economic growth to many developing countries in Latin America (Cárdenas et al. 2000; Rodrik 1998) and (East) Asia (Amsden 1989, 2001; Wade 2004). Along with GDP growth came labor productivity increases (through rapid capital accumulation), which translated into higher real wages of those employed, rising standards of living, and the development of substantial technological capabilities in manufacturing. The "revisionists" argue that ISI preceded exporting in almost all industries in almost all late-industrializing countries and in fact was a prerequisite for export-led growth, because it would not have been possible without the technological capabilities accumulated during the ISI effort that preceded it. The revisionist account emphasizes that countries pursued radically different strategies during the "mature" stage of ISI. One group of countries, including many Latin American ones, began to deepen import substitution in intermediate and capital goods. In conjunction with inappropriate monetary and fiscal policies (which led to frequent periods of high domestic inflation), this contributed to significant exchange rate instability, with large adverse effects on nontraditional exports growth and export diversification. The result was unsustainable external balances, growing international indebtedness, and collapse in response to negative external shocks (Cárdenas et al. 2000). The other group of countries, including most East Asian newly industrializing economies (NIEs), began to place substantially more emphasis on export promotion, tying incentives to export performance, while continuing with ISI. This "mixed" model of ISI became a success, unlike the "deepened" model, because NIEs' governments (1) were able to discipline private-sector firms by means of monitorable performance standards (e.g., export targets) in exchange for subsidies to make private manufacturing activities profitable (Amsden 2001); (2) avoided external imbalances by promoting exports and maintaining competitive exchange rates; and (3) succeeded in designing protection and export-promotion measures that induced indigenous technological learning and knowledge accumulation (Bruton 1998). From the revisionist literature follow important, still unresolved political-economy ques-

tions related to the nature of government, the main one of which is how NIEs' governments could induce the technological learning in private-sector firms that is usually assumed to be the consequence of market competition.

See also dependency theory; export promotion; infant industry argument; linkages, backward and forward; trade and economic development, international

FURTHER READING

Amsden, Alice H. 1989. *Asia's Next Giant: South Korea and Late Industrialization*. New York: Oxford University Press. Informative examination of Korea's industrialization in both a comparative and historical context, which challenges much conventional economic thinking on development.

———. 2001. *The Rise of "The Rest": Challenges to the West from Late Industrializing Economies.* Oxford: Oxford University Press. A masterful analysis of major late industrializing countries in which the accumulation of industrial knowledge based assets assumes center stage.

Balassa, Bela, Joel Bergsman, Pedro Malan, Teresa Jean neret, Gerardo Bueno, John Power, Stephen Lewis, Stephen Guisinger, and Preben Munthe. 1971. *The Structure of Protection in Developing Countries*. Balti more: Johns Hopkins University Press. Influential World Bank study of the welfare impact of protection in seven developing countries.

Bhagwati, Jagdish N. 1958. "Immiserizing Growth: A Geometric Note." *Review of Economic Studies* 25 (3): 201 5. Examines circumstances in which declines in the (barter) terms of trade outweigh the welfare benefits of trade induced growth.

———. 1978. *Foreign Trade Regimes and Economic Development: Anatomy and Consequences of Exchange Control Regimes*. Cambridge, MA: Ballinger. First synthesis volume of an influential NBER project on trade policy regimes and development, highlighting the downside of ISI.

———. 1982. "Directly Unproductive, Profit seeking (DUP) Activities." *Journal of Political Economy* 90 (5): 988 1002. Pioneering contribution to the political economy of (endogenous) trade policy.

Bruton, Henry J. 1998. "A Reconsideration of Import Substitution." *Journal of Economic Literature* 36 (2):

903 36. Balanced evaluation of ISI, which emphasizes technological learning and knowledge accumulation as the prime sources of growth.

Cardénas, Enrique, José Antonio Ocampo, and Rosemary Thorp. 2000. *An Economic History of Twentieth Century Latin America. Industrialization and the State in Latin America: The Postwar Years*. London: Palgrave Mac millan. Useful essays on Latin American state led in dustrialization from the early 1940s to the 1990s.

Chang, Ha Joon. 2002. *Kicking Away the Ladder. Development Strategies in Historical Perspective*. London: An them Press. An original historical analysis of the interventionist trade and industrial policies followed by the now developed countries during the 19th century.

Edwards, Sebastian. 1993. "Openness, Trade Liberalization and Growth in Developing Countries." *Journal of Economic Literature* 31 (3): 1358 93. Provides a comprehensive review of the literature up to the early 1990s.

Gerschenkron, Alexander. 1962. *Economic Backwardness in Historical Perspective*. Cambridge, MA: Harvard University Press. Pioneering and very influential formulation of the specific problems facing late industrializing nations.

Krueger, Anne O. 1978. *Foreign Trade Regimes and Economic Development: Liberalization Attempts and Consequences*. Cambridge, MA: Ballinger. Second synthesis volume of the NBER project on trade policy regimes and development, which supports export orientation and reliance on price incentives rather than ISI policies.

List, Friedrich. 1909 [1841]. *The National System of Political Economy*. Book 4. Translated by Sampson S. Lloyd. London: Longman, Green. Original protest against the principle of absolute free trade, emphasizing the impor tance of interventionist (ISI) policies for development.

Little, I.M.D., T. Scitovsky, and M. Scott. 1970. *Industry and Trade in Some Developing Countries: A Comparative Study.* New York: Oxford University Press. One of the first studies to estimate the effective rates of protection associated with ISI and to highlight negative develop mental effects.

Manoïlesco, Mihaïl. 1928. *Théorie du Protectionisme*. Paris: Marcel Giaud. Intense advocacy of ISI policies, which was further developed by Raúl Prebisch.

Nurkse, Ragnar. 1959. *Patterns of Trade and Development*. Stockholm: Almquist and Wicksell. Masterful combi

nation of history and economic theory arguing that trade could not be an engine of growth for post World War II late industrializing countries.

Ocampo, José Antonio, and Lance Taylor. 1998. "Trade Liberalization in Developing Countries: Modest Benefits but Problems with Productivity Growth, Macro Prices and Income Distribution." *Economic Journal* 108 (453): 1523 46. Substantive critique of the argument that trade is good for growth from a modern structuralist perspective.

Prebisch, Raúl. 1950. *The Economic Development of Latin America and Its Principal Problems.* New York: United Nations. A classic treatise that argues for ISI as a means to escape from poverty and primary production specialization.

Rodríguez, Francisco, and Dani Rodrik. 2001. "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross National Evidence." In *NBER Macroeconomics Annual 2000*, edited by Ben S. Bernanke and Kenneth Rogoff. Cambridge: MIT Press, 261 25. Offers a well argued and empirically grounded critique of major studies investigating the link between trade policy and growth.

Rodrik, Dani. 1998. "Globalization, Social Conflict, and Economic Growth." *World Economy* 21 (2): 143 58. Well argued defense of ISI policies.

———. 1999. *Making Openness Work: The New Global Economy and the Developing Countries.* Washington, DC: The Overseas Development Council. Careful analysis of the negative effects on development of openness to trade and finance, highlighting the scope for government action in promoting development under globalization.

Srinivasan, T. N., and Jagdish Bhagwati. 2001. "Outward Orientation and Development: Are Revisionists Right?" In *Trade, Development and Political Economy: Essays in Honor of Anne O. Krueger*, edited by D. Lal and R. H. Snape. New York: Palgrave, 3 26. Substantive review of the cross country regression studies on trade policy and growth in defense of the virtues of trade openness.

Storm, Servaas, and C.W.M. Naastepad. 2005. "Strategic Factors in Economic Development: East Asian Industrialization 1950 2003." *Development and Change* 36 (6): 1059 94. Useful review of the role of ISI in East Asian economic development.

Wade, Robert. 2004. *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization.* Princeton, NJ: Princeton University Press. A skillful analytical account of Taiwan's governed economic development in which ISI played a key part.

Waterbury, John. 1999. "The Long Gestation and Brief Triumph of Import substituting Industrialization." *World Development* 27 (2): 323 41. Examines the highly diverse origins of what became known as ISI.

Winters, L. Alan. 2004. "Trade Liberalization and Economic Performance: An Overview." *Economic Journal* 114 (February): F4 F21. Useful survey emphasizing the serious methodological problems that arise when trying to assess the impact of trade policy on economic performance.

SERVAAS STORM

■ impossible trinity

The impossible trinity describes an apparently inescapable constraint faced by policymakers, namely the impossibility of simultaneously fixing an exchange rate, permitting full capital mobility, and implementing a monetary policy with only domestic goals in mind. A fixed exchange rate implies, by the law of purchasing power parity (PPP), an inexorable link between domestic and foreign price levels. A less restrictive interpretation of PPP implies that domestic inflation is completely determined by foreign inflation. Clearly, if this is the case then a monetary policy geared toward purely domestic objectives cannot succeed. Similarly, a fixed exchange rate that operates in a world where capital is perfectly mobile must lead to a loss of control over domestic monetary policy.

Imagine an exchange rate that is undervalued that is, the domestic price of the foreign currency (i.e., the exchange rate) is above what is perceived to be the equilibrium exchange rate. If interest rates happen to be the same in the two countries then, other things being equal, foreign investors will find domestic investments more attractive since, assuming the present situation does not persist, and the exchange rate returns to equilibrium, in the future

foreign investors will be able to buy more of the foreign currency when their investment matures. As a result, the mere fixing of an exchange rate out of equilibrium creates forces that influence domestic interest rates. The important point is that a fixed exchange rate and full capital mobility can be incompatible with an independently determined domestic monetary policy (Obstfeld and Rogoff 1995).

Some economists have argued that the unwillingness to accept the existence of this trinity led many emerging economies (middle-income countries with some access to global financial markets) to adopt "soft" pegs instead of the "hard" peg assumed in the foregoing description of the impossible trinity phenomenon (Calvo and Reinhart 2002). Essentially, a hard peg establishes a rigidly fixed exchange rate that leaves relatively little room for change in the event of some unexpected event. In contrast, a soft peg is a fixed exchange rate system where the policy authorities, either explicitly or implicitly, leave room for adjustments or changes in the exchange rate as circumstances warrant. Others have pointed out that the existence of an impossible trinity is confirmed by the failure of the so-called soft pegs as demonstrated by the currency crises in emerging markets, such as Mexico in 1994; Thailand, Indonesia, and Korea in 1997; Russia and Brazil in 1998; and Argentina and Turkey in 2000 (Fischer 2001). The consequence is that exchange rate systems throughout the world have become bipolar in nature: they gravitate toward fixed or floating exchange rates. Intermediate exchange rate regimes are, therefore, a vanishing breed.

Not all interested observers agree with the notion that the impossible trinity is an immutable economic force. An understanding of these different positions requires an exploration of the proposition of the so-called vanishing middle.

The Vanishing Middle? The vanishing middle phenomenon is a corollary of the impossible trinity. Since it not possible to simultaneously fix an exchange rate, permit full capital mobility, and implement a monetary policy with only domestic goals in mind, economies will either gravitate toward a pegged exchange rate or permit their exchange rate to freely float. In order to understand the significance of

the impossible trinity it is necessary to briefly examine how exchange rate classifications are made. Until approximately 2000, the long-standing practice of the International Monetary Fund (IMF) was to report exchange rate regimes according to each member country's own view of how its exchange rate is set. As such, the resulting classification can be likened to a type of de jure classification.

By 2000, the IMF was confronting the fact that de jure and de facto definitions of the exchange rate need not be the same. Several academics (e.g., Levy-Yeyati and Sturzenegger 2005) were beginning to modify official classifications. They were dissatisfied with how governments interpreted the type of exchange rate regime they had in place when they reported their view to the IMF. However, the classification used by the economist Fischer (2001) in making the case for the impossible trinity continued to distinguish among pegged rates in a horizontal band, a crawling peg, and managed floating regimes with no preannounced exchange rate path. This suggests that a managed float differs from certain types of pegged exchange rate regimes only because the desired exchange rate is not preannounced. However, if one were also to classify exchange rate regimes based on their credibility, transparency, or durability, then it is difficult to see what sets these types of exchange rate regimes apart from one another. The soft versus hard peg classification permits a more subtle distinction between types of pegged exchange rate regimes though again, based on the historical record, or using the metric of central bank transparency, the distinction is more a matter of convenience than one with meaningful economic implications. In general, whether central banks manage international reserves owned by the government, or own them outright, they are almost always responsible for initiating intervention.

The economists Reinhart and Rogoff (2004) take a different approach to setting up an exchange rate regime classification scheme according to five categories: pegged, limited flexibility, managed float, freely floating, and freely falling. The authors conclude that far fewer countries than previously thought actually chose either one of the corner so-

lutions of the peg and the free float. More important is their finding that floating regimes deliver relatively lower inflation, but possibly at the cost of a lower average per capita economic growth rate than less flexible exchange rate regimes. The point is that the difficulty of defining a true peg suggests, as noted earlier, that a variety of managed floating regimes exists. Nevertheless, classifications still rely on whether the central bank and/or the government have a particular exchange rate regime in mind and how much variability is permitted in the exchange rate. No consideration is given to the possibility that an exchange rate regime in which intervention is used to influence uncertainty in exchange rate movements is, in some sense, also a type of managed regime.

Also absent from the discussion so far is whether it makes any difference whether the economies in question are advanced, emerging, or developing. It could be argued that one's interpretation of whether in fact exchange rate regimes are evolving to one corner of the exchange rate regime type spectrum or the other, namely, whether countries are increasingly adopting either fixed or floating regimes, is strongly related to restrictions placed on capital movements (Eichengreen and Razo-Garcia 2006). More advanced countries are much more likely to permit the free movement of capital than developing countries and, therefore, are more likely to adopt a floating exchange rate regime.

Can a Country Escape the Impossible Trinity?
The concept of a trinity immediately brings to mind the notion of an inescapable constraint in the menu of policies available to sovereign nations. Of course, the arguments are expressed in purely economic terms. However, when political dimensions are considered the policy choice dilemma facing the authorities can become even more complicated. After all, governments typically choose the exchange rate regime while the monetary policy strategy, at least in short run, is often the responsibility of the central bank. Rodrik (2000) refers to a "political trilemma," which is a variant on the impossible trinity concept. The trilemma seeks to reconcile the influence of economic integration on the nation-state. The latter

is arguably less "independent" the greater the influence of globalization. Whereas globalization involves the internationalization of economic forces, international agreements such as the Bretton Woods arrangement after World War II link nation-states through formal international political structures. Finally, at the other end of the trilemma are attempts to create a global form of federalism wherein nation-states relax the divisions created by national borders for some greater political or economic goal. The European Union immediately comes to mind.

Perhaps the best way to summarize the state of the debate is to quote from Cooper (1999, 16 17):

What is less obvious is that floating rates, independent monetary policy and freedom of capital movement may also be incompatible, at least for countries with small and poorly developed capital markets, i.e., for most countries. That would leave a more limited menu of choice for such countries: between floating rates with capital account restrictions and some monetary autonomy, or fixed rates free of capital restrictions but with loss of monetary autonomy. Put bluntly, two prescriptions regularly extended to developing countries by the international community, including the IMF and the US Treasury, namely to move toward greater exchange rate flexibility and to liberalize international capital movements, may be in deep tension, even deep contradiction.

See also band, basket, and crawl (BBC); Bretton Woods system; capital mobility; carry trade; currency board arrangement (CBA); currency crisis; discipline; European Monetary Union; exchange rate regimes; exchange rate volatility; fear of floating; interest parity conditions; International Monetary Fund (IMF); international reserves; purchasing power parity

FURTHER READING

Calvo, Guillermo, and Carmen Reinhart. 2002. "Fear of Floating." *Quarterly Journal of Economics* 117: 379 408. A technical article that makes the case that countries that act as if they have a floating exchange rate actually try to manipulate exchange rate movements.

Cooper, Richard. 1999. "Exchange Rate Choices." Harvard Institute for Economic Research Discussion Paper No. 1877 (July). Cambridge, MA: Harvard University. Summarizes the views, pro and con, of the choice be tween a fixed or flexible exchange rate regime.

Eichengreen, Barry, and Raul Razo Garcia. 2006. "The International Monetary System in the Last and Next 20 Years." *Economic Policy* 47 (July): 393 442. Written for the well informed general reader and provides an over view of the difficulties and consequences of alternative ways of defining different types of exchange rate regimes.

Fischer, Stanley. 2001. "Exchange Rate Regimes: Is the Bipolar View Correct?" *Journal of Economic Perspectives* 15: 3 24. Argues that countries around the world are choosing either a peg or a pure float.

Levi Yeyati, Eduardo, and Federico Sturzenegger. 2005. "Classifying Deeds versus Words." *European Economic Review* 49 (August): 1603 35. Proposes a new way of classifying exchange rate regimes for a large number of countries around the world. More than one classification type is proposed.

Obstfeld, Maurice, and Kenneth S. Rogoff. 1995. "The Mirage of Fixed Exchange Rates." *Journal of Economic Perspectives* 9 (fall): 73 96. Fairly accessible to readers with some economics training that essentially amounts to a critique of fixed exchange rates.

Reinhart, Carmen, and Kenneth S. Rogoff. 2004. "The Modern History of Exchange Rate Arrangements: A Reinterpretation." *Quarterly Journal of Economics* 199: 1 48. An alternative classification to the one by Levy Yeyati and Sturzenegger, based on an exhaustive analysis of exchange rate policies in a large sample of countries around the world and over the past few decades.

Rodrik, Dani. 2000. "How Far Will International Eco nomic Integration Go?" *Journal of Economic Perspectives* 14 (winter): 177 86. Ties in the phenomenon of glob alization with the constraints implied by the impossible trinity.

PIERRE L. SIKLOS

■ infant industry argument

The infant industry argument suggests that an industry may be developed under the umbrella of the government's temporary protection. Such a policy must weigh the future cost savings of an industry in which dynamic economies of scale are present with the current consumers' foregone rents due to higher domestic prices as well as higher imported prices of similar products. By establishing an import tariff that is somewhat related to the efficiency level of the domestic industry relative to the foreign one, the government articulates a rent redistribution mechanism, from domestic consumers to local producers, that may help the local industry to overcome the initial cost disadvantage and thus survive in the long run.

The argument that local industry can develop only if given a chance to reduce costs has been around for a long time. During the first Washington administration (1788 92), the first U.S. secretary of the treasury, Alexander Hamilton, favored the temporary protection of the American industry to facilitate its full and fast development as an effective way for the United States to become less dependent on English manufactures. Import substitution can certainly be welfare enhancing if learning is fast enough so that near-future cost savings overcome current consumer losses. It should be pointed out that an import duty not only makes imports more expensive. It also allows domestic producers of similar products to charge higher prices, thus softening domestic competition. This is perhaps the reason why Thomas Jefferson, then secretary of state of the Washington administration and representative of the rural South, bitterly opposed Hamilton's protection plan. Import duties immediately made consumption of all products more expensive for southerners while the North reaped the gains through an increase in jobs and industrial capability. Hamilton's policy as later Friedrich List's proposals for Germany was intended to be temporary but it was still around more than a half century after it was first implemented, and the issue of protection was second only to slavery in the contentious relations between the North and the South in the period leading up to the Civil War.

The academic debate surrounding the infant industry argument focuses on whether this policy can be effective and on analyzing whether the future gains

offset current costs. But this is a debate that is not alien to the influence of political ideology. Left-leaning parties traditionally favor the involvement of governments in economic activity and as a consequence they traditionally favor a protectionist policy based on the infant industry or any other argument. Conservative parties favor promarket policies, lately favor free-trade agreements, and commonly have denounced the infant industry argument based on failed experiences or have pointed out how easily a projected temporary protection can become permanent.

Infant industry tariff protection was successfully implemented in Japan after World War II, establishing the principle that has inspired the import-substitution policy of many Latin American and Asian countries for many decades with varying degrees of success. Despite many promises of an imminent opening to trade, successive governments repeatedly renewed tariff protection. Japan succeeded in developing its industry and some other countries later moved toward a policy of export promotion in order to ensure fast learning and the development of the domestic industry. The Latin American experience was less positive. Domestic industries failed to take off despite repeated renewals of protection, while locals felt impoverished as imports and related domestic products became more expensive. Furthermore, as domestic producers felt that their respective governments would renew protection, they had little interest in innovating and becoming more efficient. If the dynamic gains of protection are never realized because domestic producers do not believe that protection is temporary, the effects of the infant-industry protection can be analyzed within a static setup with a domestic industry with market power. In the case of Latin America a well-intended but poorly designed policy led to a massive transfer of rents from consumers to producers, who saw their market power significantly increased and the incentives to innovate reduced.

As noted earlier, an import duty does more than just increase the price of imports. Local producers of similar products can charge higher markups as foreign competitors lose at least part of their cost advantage. These are not the only costs of protection. In addition, protection induces a loss of competitiveness of industries that use the output of the infant industry as inputs. These "collateral costs" of protection are ignored in the partial equilibrium analysis commonly used to evaluate the welfare implications of the infant industry argument. However, since governments commonly fail to lift protection, these additional costs of establishing a "temporary" import duty confirm the current negative opinion that the overwhelming majority of economists share regarding any tariff protection based on potential learning effects.

The temporal trade-off of cost and benefits motivates the research question surrounding the infant industry argument. Could the infant industry have been developed without protection? If so, what is the cost of this development? Will the infant industry ever be able to compete with developed foreign firms? In welfare terms, can protection lead to a higher total surplus than simply allowing imports at competitive prices? And last, will protection ever end or is the infant industry argument just another excuse to prolong protection indefinitely?

Perhaps the most important criticism of the infant industry argument is that governments lack any incentive to lift such protection in the future. If that is the case, any incentive that protection may introduce for firms to invest and take advantage of this temporary protection will disappear, and thus the country will end up, as many undeveloped economies nowadays, with high import duties but without dynamic industrial development.

The practical implementation of the infant industry argument is certainly difficult. Governments must be able to accurately predict the learning dynamics of each particular industry to be protected as well as the cost that such protection generates to domestic consumers and firms. From a theoretical point of view, the question is a different one: Is infant industry tariff protection "logically" time-inconsistent? Should we oppose such protection on the basis of a theoretical result or just because from a practical perspective we believe that its effectiveness is quite implausible?

A majority of economists have argued against infant industry protection mostly from ideological premises rather than based on an appropriate theoretical model. Thus the infant industry argument is frequently dismissed by pointing out that free trade is Pareto dominant, that is, it would be mutually beneficial to all countries involved. Neither Hamilton nor the protectionists ever claimed such a global viewpoint. The infant industry argument therefore needs to be addressed within the partial equilibrium framework of a small country and evaluated on the basis of whether a temporary protection policy can effectively help develop a domestic infant industry that otherwise would never have taken off.

The academic debate on infant industry protection has recently focused on precisely the issue of whether tariff protection will ever end after successfully helping the domestic industry to develop. To summarize the results described in the rest of this entry: infant industry protection may be effective and temporary when the level of protection is linked to the efficiency level of the domestic industry. However, it will commonly fail if it sets a future liberalization date that is independent of the degree of development of the domestic industry, or if cost reduction requires specific investments that are not directly linked to the production decisions of domestic firms.

A Framework for the Analysis of Infant Industry Protection Consider a small country where an early industry suffers from such cost disadvantage relative to foreign producers that it will have to shut down in the event that the domestic government enforces a free trade policy. The domestic and foreign industries produce similar, but not necessarily identical products. The foreign industry is assumed to behave competitively and to have exhausted all its dynamic economies of scale. A single infant monopolist is assumed to produce all domestic production. This firm enjoys significant learning by doing and marginal cost will fall as production takes place. The demand for differentiated domestic and imported products depends on the decision variables of the two players: government chooses the import tariff to maximize the discounted sum of consumer surplus, profits, and tax revenues while the monopolist chooses the relative price of the domestic good that maximizes the net present value of future profits. Marginal cost reduces at a certain rate with current production and experience depreciates as a fraction of the current level of marginal cost. Therefore, learning is a reversible process that requires some positive output to induce cost savings while at the stationary equilibrium, once learning is exhausted, current production only impedes marginal cost from increasing.

The monopolist and the government engage in a dynamic game where the former chooses the price and the latter the tariff level. Both players discount the future at a common rate. The game is dynamic because each player's action affects the state of the game, that is, the level of marginal cost through the direct or indirect effect on the domestic production and accumulation of experience. This setup is essentially a capital accumulation game World such as the one studied by Reynolds (1987). If we further restrict our attention to a system of linear demands for domestic products and imports, this dynamic model becomes a linear-quadratic differential game that allows us to obtain a closed-form solution and thus easily characterize the features of the equilibrium strategies.

Alternative Assumptions The foregoing description is based on Miravete (2003). Melitz (2005) departs from this framework in two main aspects: the domestic industry is assumed to be competitive and the government behaves as a social planner.

The earlier framework could be easily generalized to an *n*-firm symmetric oligopoly industry without changing qualitative results other than learning would now be slower as each firm would only count for one *n*th of accumulated experience. Thus the more firms in the domestic industry, the faster learning needs to be in order to achieve higher welfare than the default scenario where domestic firms just shut down and consumers only purchase from foreign firms. An additional issue that may arise when

several firms compete in the domestic market is the existence of learning spillovers. As Stokey (1986) shows, this introduces additional reasons to protect an industry.

By assuming the existence of a social planner rather than a noncooperative solution, Melitz eliminates the dynamic interaction between government and firms. This corresponds to a situation in the above framework when the government and the monopolist collude in choosing the tariff and price strategies, which may be important for state-owned industries or when the industry in question has significant political influence. In a dynamic strategic environment firms would charge higher prices and the government lower rates than in Melitz's case.

This framework presents another important advantage: an infinite horizon game rules out the possibility of any other future rent after the formal end of the game, and the equilibrium strategies are thus robust to the existence of any other unaccounted rent. This is not the case of Leahy and Neary (1999), Miyagiwa and Ohno (1999), and most explicitly Tornell (1991), who considers the possibility of a third period where wages are renegotiated when the initial protection of the industry was planned only for two periods. Considering extraneous elements after the planning horizon of the game as formally ended turns protection, by definition, into a time-inconsistent policy.

Markov versus Non-Markov Strategies The basic criticism to the infant industry argument is that governments lack the ability to commit to taking particular policy actions in the future. In particular, a government cannot commit to reducing protection if the protected industry fails to make the anticipated productivity gains. Although this may be true in practice, it cannot be defended with models that by construction make reneging from the announced policy the dominant government's strategy. A more interesting distinction is the use of strategies that are linked to the level of efficiency of the domestic industry. The effect of learning is reducing the marginal cost of production simply by means of producing in the past. A protection strategy linked to this marginal cost (or alternatively to the related level of output or price) of the domestic industry is called a Markov strategy.

A way to help governments renege from announced liberalization is by using non-Markov strategies such as in the case of Staiger and Tabellini (1987) or Matsuyama (1990). Protection eases domestic production, and thus, through learning, marginal costs get reduced. In these models, however, cost reduction depends on investments that are not linked to the state of the game, that is, the level of marginal cost. Therefore there is no source of commitment for the government either. In the framework described earlier such policies can indeed be computed when both the government and the domestic producer can commit to a sequence of tariffs and prices over time. But most generally, a Markov perfect equilibrium can be constructed when both tariffs and prices are made contingent on the evolving level of marginal cost. The interesting added result of taking this approach is that if such strategies are employed, the best response of the government to the pricing of firms is to reduce tariff protection as the level of marginal cost decreases. Thus there is no need to renege and time-consistent equilibrium strategies lead to future trade liberalization. The government simply allows the domestic firm to earn the minimum markup to induce them to reduce costs beyond the static equilibrium. As this process takes place, marginal cost gets reduced, and the required level of protection needed is also lower. But this result also confirms Hamilton's intuition, that protection based on the infant industry argument would be only temporary despite sometimes surviving for a long time.

See also import substitution industrialization; trade and economic development, international

FURTHER READING

Leahy, Dermot, and Peter J. Neary. 1999. "Learning by Doing, Precommitment and Infant industry Promotion." *Review of Economic Studies* 66 (2): 447 74. A model of tariff protection with a finite horizon.

Matsuyama, Kiminori. 1990. "Perfect Equilibria in a Trade Liberalization Game." *American Economic Review* 80

(3): 480 92. A game theoretical model that shows that permanent protection may arise from the lack of commitment of tariff protection policies.

Melitz, Marc J. 2005. "When and How Should Infant Industries Be Protected?" *Journal of International Economics* 66 (1): 177 96. An elegant model of learning by doing that points out how the speed of learning determines the length of protection needed by an industry.

Miravete, Eugenio J. 2003. "Time consistent Protection with Learning by Doing." *European Economic Review* 47 (5): 761 90. A differential game model that shows that time consistent protection is possible in an infinite horizon differential game where firms' and the government's strategies are contingent on the level of marginal cost of the industry.

Miyagiwa, Kaz, and Yuka Ohno. 1999. "Credibility of Protection and Incentives to Innovate." *International Economic Review* 40 (1): 143 63. A dynamic model of industry protection with policies that are not linked to industry performance.

Reynolds, Stanley S. 1987. "Capacity Investment, Preemption, and Commitment in an Infinite Horizon Model." *International Economic Review* 28 (1): 69 88. A differential game theoretical framework that studies capital accumulation games.

Staiger, Robert W., and Guido Tabellini. 1987. "Discretionary Trade Policy and Excessive Protection." *American Economic Review* 77 (5): 823 37. A model of non Markov protection strategies that shows how the lack of commitment leads to high tariffs.

Stokey, Nancy L. 1986. "The Dynamics of Industrywide Learning." In *Equilibrium Analysis: Essays in Honor of Kenneth J. Arrow*, vol. 2, edited by W. P. Heller, R. M. Starr, and D. A. Starret. New York: Cambridge University Press, 81 104. An elegant dynamic model of industry performance where learning occurs in the early stages of development through spillover effects.

Tornell, Aaron. 1991. "Time Inconsistency of Protectionist Programs." *Quarterly Journal of Economics* 106 (3): 963 74. A model of finite horizon protection where time inconsistency arises by extending the length of the game once the strategies have been defined.

EUGENIO J. MIRAVETE

■ inflation targeting

Since the early 1990s, inflation targeting has become the predominant monetary policy choice for central banks in industrial and developing economies alike. New Zealand, the United Kingdom, Canada, Australia, Norway, Sweden, and Israel were among the first countries to adopt this policy. Since then, many emerging market economies have also adopted it, including Chile, Brazil, South Africa, Korea, Indonesia, Thailand, and the Philippines. Inflation targeting is currently regarded as best practice monetary policy for countries operating flexible exchange rate regimes.

At its core, inflation targeting is a monetary policy system that manipulates a policy instrument (usually a short-term nominal interest rate, but sometimes the nominal effective exchange rate) according to available information with the objective of controlling inflation. The rate of inflation is the primary objective of this regime but is not the sole objective; the system permits the pursuit of secondary objectives such as smoothing changes in output, interest rates, or even the exchange rate.

The successful implementation of inflation targeting depends on a checklist of features:

1. A binding commitment to price stability as the primary policy objective. This is normally presented in a central bank act or similar piece of legislation that sets the objectives and operational rules for the functioning of a central bank.
2. Transparency. This takes the form of announced inflation targets, published minutes of meetings of the monetary policy committee, and the publication of periodic inflation reports.
3. An independent and accountable central bank.
4. A floating exchange rate. In a world of high capital mobility, having an exchange rate target compromises the efficacy of inflation targeting. Nonetheless, central banks in emerging markets often have a strong desire to manage exchange rate movements while maintaining an inflation targeting regime.

5. A monetary policy rule (MPR). The MPR is a policy rule that guides the instrument of the inflation targeting policy. If the nominal interest rate is the policy instrument, the MPR states how it should react to key economic variables such that the policy objective is attained (see Taylor 2001).

Inflation targeting has several advantages. First, it enables policymakers to focus on domestic policy considerations and to respond to those shocks affecting the domestic economy. Second, inflation targeting does not require a stable relationship between the money supply and inflation but uses all available information deemed useful in achieving the target. Third, inflation targeting is transparent and easily understood by the public (Mishkin 2000). Inflation targeting also has two notable disadvantages. The first is that, owing to the long and variable lags between the time a policy is instigated and point it begins to take effect, inflation is inherently difficult to control. The second, which is pertinent to emerging market economies, is that the exchange rate flexibility associated with inflation targeting may lead to financial instability.

A strict inflation target provides that inflation is the *sole* objective of monetary policy. In other words the only variable the central bank cares about is inflation. Flexible inflation targeting essentially targets inflation but allows for a secondary objective. The usual secondary objective is output, where the central bank seeks to minimize the gap between the current growth rate of output and some steady-state value. Other secondary objectives may include dampening the growth of asset prices, reducing the volatility of the exchange rate, or minimizing potential disruptions in domestic financial markets caused by sharp changes in interest rates. It is commonly held and demonstrated in simulation studies (see Svensson 1997) that output variability is lower under flexible inflation targeting and that the inflation target is attained more quickly under strict inflation targeting. Central banks are often under pressure to take output conditions into account, however, and in this situation, flexible inflation targeting is a viable option.

There are several other analytical and operational dimensions to inflation targeting that are significant in its construction. These include the central bank's selection of a strict versus flexible inflation target, whether it should target domestic inflation (inflation in the price of those goods produced and consumed domestically) or consumer price index inflation (domestic inflation plus inflation in the price of tradable goods), whether the central bank should target current inflation or an inflation forecast, and the extent to which the central bank should respond to the exchange rate and asset prices.

See also exchange rate regimes; monetary policy rules; money supply; seigniorage

FURTHER READING

Ball, L. 1997. "Efficient Rules for Monetary Policy." NBER Working Paper No. 5952. Cambridge, MA: National Bureau of Economic Research. Presents the basic analytics for a closed economy strict and flexible inflation targeting system and the specification of the rule(s) required to reach the inflation objective.

Cavoli, T., and R. Rajan. 2006. "Inflation Targeting in East Asia: Exploring the Role of the Exchange Rate." *Briefing Notes in Economics*, No. 74 (September October). Similar to Ball (1997) but for open economies where the role of the exchange rate as a policy objective is examined.

Debelle, G. 2001. "The Case for Inflation Targeting in East Asian Countries." In *Future Directions for Monetary Policy in East Asia*, edited by D. Gruen. Canberra: Reserve Bank of Australia, 65 82. Examines the issues regarding the appropriateness of inflation targeting in developing economies particularly in Asia.

Mishkin, F. S. 2000. "Inflation Targeting in Emerging Market Countries." *American Economic Review Papers and Proceedings* 90 (2): 105 9. Evaluates the checklist for inflation targeting for developing economies.

Svensson, L.E.O. 1997. "Inflation Forecast Targeting: Implementing and Monitoring Inflation Targets." *European Economic Review* 41 (6): 1111 46. Presents a technical treatment of the trade offs inherent in strict versus flexible inflation targeting and examines whether central banks should target actual current inflation or a forecast of future inflation.

Taylor, J. B. 2001. "The Role of the Exchange Rate in Monetary Policy Rules." *American Economic Review Papers and Proceedings* 91: 263 67. Examines the issues of fitting the exchange rate in an inflation targeting framework.

TONY CAVOLI

■ information and communications technology

Information and communications technology (ICT) refers to all technologies and devices used in managing and processing information systems. In contrast to the manufacturing industries such as shipbuilding, steel, automobiles, and textiles, which create value directly, ICT in the form of computers, software, the Internet, multimedia, and management of information services creates value indirectly. ICT includes data for business use, voice communication, images, multimedia, and many types of technologies for the purpose of development and exchange of information. The definition of the ICT sector differs among industrialized countries. The U.S. and Organisation for Economic Co-operation and Development (OECD) definitions are presented in table 1.

The introduction of increased processing power of hardware along with complementary software and telecommunications infrastructure have enhanced the ability to store, retrieve, analyze, and communicate data and information within and between organizations and their partners and suppliers, and ultimately to the consumer. In particular, ICT deals with the use of electronics, computers, and computer software to convert, store, protect, process, transmit, and retrieve information. For that reason, computer professionals are often called information technology (IT) specialists or business process consultants, and the division of a company or university that deals with software technology is often called the IT department. Other names for the latter are information services, management information services, or managed service providers.

As a general rule, capital and labor as important elements in the growth of an economy lead to decreasing returns to scale. Therefore, even though these factor inputs of production may increase, the growth of an economy over a certain level cannot be expected. Information may produce increasing returns to scale and become an important factor for sustainable growth, however. Thus, in recent years, ICT has been considered an input in the production of goods and services and a factor affecting total factor productivity growth at both the micro- and the macro levels (see Shiu and Heshmati 2006).

Review of the ICT Literature ICT is considered one of the three major technological breakthroughs of the modern era (see Edquist and Henrekson 2007), the others being steam power and electricity. ICT includes some of the wider information technology innovations and applications, and their commercialization and transfer have been quite rapid. These enhance the communication of more accurate and value-added information to workers, managers, and consumers, thus reducing uncertainty and time use in conducting many types of businesses.

There has been great interest among researchers in investigating how some countries were able to take advantage of ICT to accelerate their rates of growth and productivity. In these studies one examines the contribution of IT investment to economic growth and finds that the returns on IT investment are significantly positive. In the transitional countries of Central and Eastern Europe (CEE), one observes a chronic underinvestment in telecommunications infrastructure. The result suggests that improving investment may ultimately improve the channel between aggregate investment and growth economywide (see also Zhu 1996; Madden and Savage 1998).

Return on IT differs by the country's development level. Results from intercountry studies relating IT and non-IT inputs to gross domestic product (GDP) over time suggest that for developed countries, returns from IT capital investments are positive, while returns from non-IT capital investments are not commensurate with relative factor shares. The situation is reversed for developing countries, where

Table 1
An industrial sector based definition of ICT

U.S. Department of Commerce IT producing industries	OECD ICT industry
Hardware industries	**Manufacturing**
Computers and equipment	Manufacture of office, accounting, and computing machinery
Wholesale trade of computers and equipment	Manufacture of insulated wire and cable
Retail trade of computers and equipment	Manufacture of electronic valves, tubes, and other electronic components
Calculating and office machines	Manufacture of television and radio transmit ters and apparatus for line telephony and line telegraphy
Magnetic and optical recording media	
Electron tubes	
Printed circuit boards	Manufacture of television and radio receivers, sound or video recording or reproducing apparatus, and associated goods
Semiconductors	
Passive electronic components	
Industrial instruments for measurement	Manufacture of instruments and appliances for measuring, checking, testing, navigating, and other purposes, except industrial process control equipment
Instruments for measuring electricity	
Laboratory analytical instruments	
	Manufacture of industrial process control equipment
Communications equipment industries	
Household audio and video equipment	
Telephone and telegraph equipment	**Services: Goods related**
Radio and TV communications equipment	Wholesale of machinery, equipment, and sup plies
Software/services industries	Renting of office machinery and equipment (including computers)
Computer programming	
Prepackaged software	**Services: Intangible**
Wholesale trade of software	Telecommunications
Retail trade of software	Computer and related activities
Computer integrated system design	
Computer processing, data preparation	
Information retrieval services	
Computer services management	
Computer rental and leasing	
Computer maintenance and repair	
Computer related services, NEC	
Communications services industries	
Telephone and telegraph communications	
Cable and other TV services	

Sources: U.S. Department of Commerce (2003), OECD (2000).

returns from non-IT capital are quite substantial, but those from IT capital investments are not statistically significant. The impacts of IT investment on economic growth in a cross section of countries shows that the relative contribution of IT to GDP growth in developing counties between 1980 and 1995 was less than 2 percent, compared to more than 10 percent in developed countries (see Dewan and Kraemer 2000; Pohjola 2001).

Investment in ICT and Its Diffusion Evidence on the role of ICT investment is primarily available at the macroeconomic level. It has been observed that ICT has been a very dynamic area of investment, due to the steep decline in ICT prices, which has

encouraged investment in ICT and expansion of production at the same time shifting investment away from other assets. The capital deepening that results from investment in ICT is considered an important driver of economic growth. It establishes the infrastructure for the use of ICT networks and provides productive equipment and software to businesses. Measures of ICT investment are therefore of considerable interest in examining growth performance in many countries. Investment is usually estimated by using business surveys. These surveys usually allow the total investment to be disaggregated into a number of well-defined asset groups, including ICT. There is a broad understanding in the statistical community about the definition of ICT products. The pace of investment differs widely by country. The lowest levels of the share of investment in ICT are found in low-income nations, while the highest are in high-income nations.

Generally, network products, including telephone, e-mail, Internet, computer hardware, and software, have distinct features such as network effect, critical mass, lock-in, and path dependency, which affect late takeoff in their diffusion. A positive direct network effect means a positive utility gain for consumers when the number of users operating the same system increases. For example, the first e-mail message was sent in 1969, but the adoption did not take off until 1990. Since then, Internet traffic has been doubling every year. This example raises an important question, which is when to expect a new technology to diffuse and what should be the minimum number of users (the critical mass) needed for inducing potential consumers to adopt it.

At the global level, the diffusion of key ICTs such as the Internet, mobile phones, and personal computers in accordance with the World Bank's classifications are shown in figures 1, 2, and 3. To judge from these graphs, which are based on data published in OECD (2000; 2004), United Nations Conference on Trade and Development (2003), U.S. Department of Commerce (2003), and World Bank (2006), we observed that high-income countries show a high level of diffusion. In addition, in the case of the diffusion of the Internet and personal com-
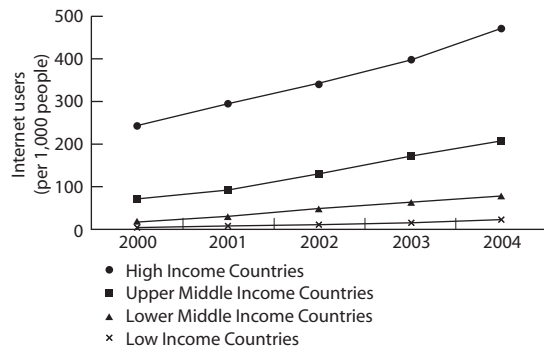
Figure 1

The diffusion of the Internet
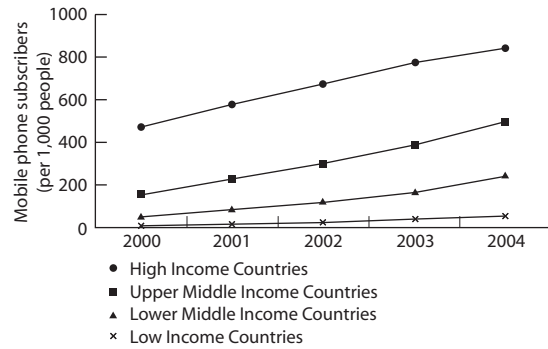
Figure 2

The diffusion of mobile phones
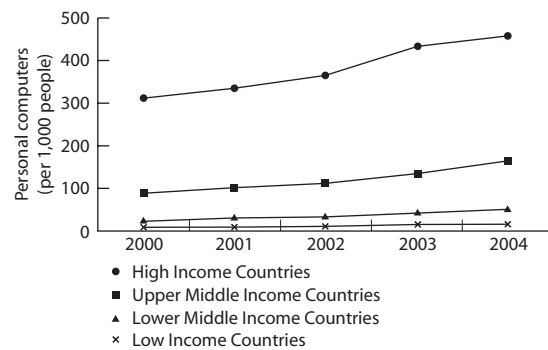
Figure 3

The diffusion of personal computers

puters, the gap between high-income and low-income countries measured in Internet users per 1,000 people increases over time.

The "digital divide" is the socioeconomic difference between communities with access to computers and the Internet and those without such access. At the microlevel, it refers to the gap between individuals, households, businesses, and geographical areas with regard to their opportunities and abilities to access and to use IT services for a wide variety of activities. The gap is due to differing literacy and technical skills, and the gap in availability of useful digital content. The digital divide at the aggregate level is often discussed in conjunction with the gap between rich and poor, and developed and underdeveloped nations concerning access to and use of digital communication. In an international context, the divide indicates that developed countries are far better equipped than developing countries to use the advantages of rapidly expanding Internet technology. The more rapidly the rate at which Internet technology is developed and spread, the more the quality-of-life differences between developed and underdeveloped countries become evident. Despite the productivity, connectivity, and many other recognized and measurable positive effects associated with Internet use, the international digital divide is widening.

Given the gradually decreasing number of fixed (land-line) telephone subscribers per 1,000 people in contrast to mobile subscribers, it is highly probable that the traditional fixed phone is being supplanted by the mobile phone as telecommunications services develop. Consumers show increasing preference for data-based communication instead of voice-based, and new communications services such as voice-over-Internet protocol have entered the market through the development of Internet technology. Given the Internet's high rate of growth and the emergence of broadband services, the difference between Internet telephony and traditional voice telephony has begun to erode. Such changes are dynamically transforming the traditional communications industry structure. The diffusion into the marketplace of telephone, mobile phone, and Internet services has

progressed differently in different places, determined mainly by local and regional economic conditions.

The Effects of ICT on Economic Growth Much attention has been given to how much IT affects the growth of an economy (see Jorgenson 2001). As is well known, IT can transform our economic system by increasing productivity and stimulating economic growth. At the firm level, there are four mechanisms or channels through which IT investment affects the growth of an economy. The first channel is that the IT industry itself grows dramatically, and the nations where the IT industry occupies a leading position may have more than one leading growth sector. In the decade between 1995 and 2005, many nations' portion of revenue derived from the IT sector in their total GDP increased. For example, even in the case of China achieving an 8 percent annual growth rate during that decade, the growth of the IT sectors was faster than the overall economic growth. Accordingly, the expansion of IT sectors affects the growth of the economy positively. By the second channel, IT can facilitate the catch-up process by activating the diffusion process of technologies. According to Antonelli (1990), developing countries can take advantage of the opportunity of catch-up by overcoming asymmetry and disequilibrium of information through the diffusion of IT. The third relation between IT and economic growth is that developed IT affects the integration and efficiency of markets and it stimulates economic growth. This is called the market integration effect. In the final mechanism, IT improves the management and decision-making processes of companies. Additionally, ICT helps firms gain market share, raise overall productivity, expand their product range, customize the services offered, respond better to client demand, and reduce inefficiency.

Economies that successfully implement new ICT may be able to overcome barriers that have long held them back in their contribution to the global trade. The rapid spread of the Internet has opened up previously unavailable commercial and political information. In particular, ICT has reduced many of the transaction costs of participating in subcontracting through business-to-business interaction, and it

facilitates the operations of low-cost suppliers of IT services based in developing countries. In other words, IT can increase total output eventually through reduced transaction costs. Enough constructed IT infrastructure decreases the cost of information, and in the long run it makes the market more effective. Less-developed countries face a higher cost to obtain information because their information market is less efficient than in developed countries.

IT investments have specific features. A decreased telecommunications cost can be used to decrease costs associated with decisions to distribute resources between cities and rural communities. The decreased telecommunications costs make the use of a larger quantity of information possible, which contributes to better decision making. It increases the opportunity for arbitrage and makes the financial market more efficient. Finally, decreased cost can provide more information about market prices. Even if ICT spread in similar ways in different places, one would not expect similar economic benefits from it. Only with a regulatory environment, the availability of appropriate skills, the ability to change organizational setups, and the strength of accompanying innovations in ICT applications can the benefits of IT be maximized. The magnitude of complementary policies for IT investment must be expanded. Such policies aim to enhance the conditions for developing the economy, through provision of infrastructure, prohibition of monopoly in the telecommunications market, allowing for new entrants, enacting efficient laws and regulations, and providing a high-quality education system.

Edquist and Henrekson (2007) examined productivity growth following the three major technological breakthroughs: the steam power revolution, electrification, and the ICT revolution. The authors distinguish between sectors producing and sectors using the new technology and find a long lag from the time of the original invention until a substantial increase in the rate of productivity growth can be observed, as well as strong evidence of rapid product price decreases over time. The highest productivity growth rates are found in the ICT-producing industries, which may be explained by the ICTs' more rapid rate of technological development compared with previous breakthroughs.

New Economy and the Productivity Paradox
The extraordinary performance of the U.S. economy in the late 1990s and the associated economic growth were referred to as the "new economy." The computing power of microchips, which underlies the rapid progress in productivity of ICT, has doubled every 18 24 months since 1965, as Moore's Law predicted. Contrary to mainstream economists' theoretical models, inflation and unemployment were low at the same time, particularly following 1992, and the economy experienced sustained growth and the stock market boomed throughout the 1990s.

The new economy phenomenon happened because of a number of factors. First, increased efficiency in firms' management though ICT adoption affected productivity growth at the firm level and connected productivity growth in each industry through spillover effects. As a consequence, productivity in the total economy increased. In other words, since the middle of the 1980s, more intensive competition between companies contributed to high investment in IT and the introduction of innovative management, which in turn affected productivity gains.

Second, productivity gains led to a low inflation rate. When productivity continues to grow, the inflation rate becomes lower because inflation is offset by productivity increases. Furthermore, low inflation coexists with low interest rates, and it increases the investment rate. Finally, it can link productivity growth and sustained economic growth in an interconnected cycle of investment, productivity, and economic growth.

Third, the wide-ranging diffusion of IT and Internet use made it possible for the new economy to evolve. The spread of IT and the Internet due to price reductions stimulated the network effect, and it induced sustained economic growth by creating increasing returns to scale in the whole economy. Furthermore, continuous strong productivity growth since the 2001 recession made it likely that some of the gains of the late 1990s may endure.

In disagreement with the widespread view about productivity gains through ICT adoption, Robert Solow commented in 1987 on the IT productivity paradox that the productivity of the workforce had not risen as IT had extended through Western industry. It was widely believed that office automation was boosting labor or total factor productivity, but the growth accounts did not seem to confirm this, because the "computer era" from the early 1970s to the time that Solow spoke about the paradox included a massive slow-down in growth as the machines were becoming ubiquitous.

The causes of the productivity paradox are the following. First, a portion of the benefit from ICT is not included in productivity statistics. Productivity statistics of service sectors with a high rate of investment in ICT correspond to this case. For example, the increased benefit of financial services such as ATMs (automated teller machines) is not included in productivity statistics. Insurance, business services, and health services show the same phenomenon of the benefits of ICT not being fully reflected in productivity statistics.

Second, there may be a lag in productivity improvements. The lag is because computers did not enhance productivity until things such as software, the Internet, and handhelds became prevalent. It takes a long time for a new technology to become popular and companies to adopt it. Given that ICT increases multifactor productivity through networks, and building a network requires a long time, productivity growth will lag behind the initial introduction of new technology. Productivity growth originating from IT may already be occurring, but a flaw in the measurement tools available may be hiding it.

Third, previously much research aimed at identifying the effect of ICT at the company level was based on small samples. If the early impact of ICT is trivial, research in the early stage will not capture the contribution of ICT. In addition, some research shows that the ICT impact differs by industrial sector. These days, there are more detailed data on ICT investment available to measure the impacts of ICT.

In the search for explanations of the productivity paradox, Oliner and Sichel (2000) deny the significance of the IT sector by arguing that IT accounted for no more than 2 percent of the capital stock in any country in the world. Other economists have made more controversial statements about the utility of computers: that they pale into insignificance as a source of productivity advantage when compared to the first industrial revolution or the adoption of motorcars.

The Role of ICT in the World Economy The OECD (2004) report on the impact of ICT provides two important messages. First, ICT continued to have a strong impact on performance. Productivity growth in the United States, the main example of ICT-led growth and productivity improvements, continued to be strong. ICT networks had spread throughout much of the OECD business sector and would spread further to enhance business performance. The release of increasingly powerful microprocessors was projected to continue for the foreseeable future. These factors were expected to encourage ICT investment and support further productivity growth.

Second, the impact of ICT differed markedly across OECD economies. Many OECD countries lagged in the diffusion of ICT and had scope for greater uptake. It was expected that the largest economic benefits of ICT would be observed in countries with high levels of ICT diffusion. Having the equipment or network is not sufficient to derive economic benefits, however. Other factors, such as the regulatory environment, skills, ability to change organizational setups, and the strength of innovations in ICT applications all affect the ability of firms to seize the benefits of ICT. Consequently, the impact of ICT on economic growth and performance of countries with equal ICT diffusion will not be the same.

IT makes a positive although small contribution to economic growth, but its impact is positively related to the level of development. Studies of the relationships between IT and economic performance suggest that the impact of IT diffusion can differ even among developed countries with similar level of

development. The limited existing empirical evidence shows that developing counties that did not adopt complementary policies have gained little benefit from IT investment. In general, for developing countries it is rather difficult to catch any systematic evidence about such relationships. While the evidence suggests that IT contributes to the growth of developed countries, this relationship is less clear in the case of developing countries.

A long time period is required to link IT investment to economic growth and establish a causal relationship between the two. For IT to be effective, its spread needs to reach a critical threshold. To obtain high returns from IT investment, developing countries must adopt complementary policies that fulfill conditions for economic development such as building up the basic infrastructure, creating a non-monopoly telecommunications market, opening the market, enacting effective laws and regulations, and building a high-quality educational system. For developing countries, diffusion and application of IT can play an important role in the growth of the economy, but for IT to have a full impact, more fundamental complementary investments are essential.

See also digital divide; electronic commerce; globalization

FURTHER READING

Antonelli, Cristiano. 1990. "Information Technology and the Derived Demand for Telecommunications Services in the Manufacturing Industry." *Information Economics and Policy* 4: 45 55. An empirical examination of the factors that influence the growth in demand for tele communication services in the Italian manufacturing sector.

Dewan, Sanjeen, and Kenneth L. Kraemer. 2000. "In formation Technology and Productivity: Evidence from Country level Data." *Management Science* 46: 548 62. An empirical analysis of significant differences between developed and developing countries with respect to their structure of returns from capital investments.

Edquist, Harald, and Magnus Henrekson. 2007. "Tech nological Breakthroughs and Productivity Growth." *Research in Economic History* 24: 1 53. An examination of productivity growth following three major techno logical breakthroughs: the steam power revolution, electrification, and the ICT revolution.

Jorgenson, Dale W. 2001. "Information Technology and the U.S. Economy." *American Economic Review* 91 (1): 1 32. An important study of the link between IT and the fundamental change in the U.S. economy leading to a permanent improvement in the U.S. growth pros pects.

Madden, Gary, and Scott J. Savage. 1998. "CEE Tele communications Investment and Economic Growth." *Information Economics and Policy* 10: 173 95. An em pirical examination of the relationship between gross fixed investment, telecommunications infrastructure investments, and economic growth.

Oliner, Stephen D., and Daniel E. Sichel. 2000. "The Re surgence of Growth in the Late 1990s: Is Information Technology the Story." *Journal of Economic Perspectives* 14 (4): 3 22. An empirical examination of the influence of ICT on productivity growth.

Organisation for Economic Co operation and Develop ment (OECD). 2000. *Information Technology Outlook 2000.* Paris: OECD. A thorough review of trends and prospects in the ICT industry.

———. 2004. *The Economic Impact of ICT: Measurement, Evidence, and Implication.* Paris: OECD. A compre hensive overview of the impacts of ICT on economic performance, and the ways through which these impacts can be measured.

Pohjola, Matti. 2001. "Information Technology and Eco nomic Growth: A Cross Country Analysis." In *Infor mation Technology, Productivity, and Economic Growth: International Evidence and Implications for Economic Development*, edited by Matti Pohjola. Oxford: Oxford University Press, 242 56. An exploration of impacts of information technology investment on economic growth in a cross section of countries.

Shiu, Alice, and Almas Heshmati. 2006. "Technical Change and Total Factor Productivity for Chinese Province: A Panel Data Analysis." IZA Discussion Paper 2133. Bonn: The Institute for the Study of Labor. An empirical analysis of the variables that affect total factor produc tivity growth in China.

United Nations Conference on Trade and Development. 2003. *E Commerce and Development Report 2003.* New York: UNCTAD. A thorough overview of the implica

tions of the growth of digital economy for developing countries.

U.S. Department of Commerce. 2003. *Digital Economy 2003*. Washington, DC: U.S. Department of Commerce. A comprehensive review of conditions in U.S. information technology industries and the effects of IT on national economic performance.

World Bank. 2006. *World Development Indicators 2006*. Washington, DC: World Bank. World Bank's premier annual compilation of data about development, including more than 900 indicators in over 80 tables organized in 6 sections: World View, People, Environment, Economy, States and Markets, and Global Links.

Zhu, Ji. 1996. "Comparing the Effects of Mass Media and Telecommunications on Economic Development: A Pooled Time Series Analysis." *Gazette* 57: 17 28. An empirical study of the contribution of telecommunications to economic development.

ALMAS HESHMATI AND MINKYU LEE

■ infrastructure and foreign direct investment

Foreign direct investment (FDI), acquiring a controlling share of a foreign firm, is more likely to occur in countries with good physical infrastructure such as bridges, ports, and highways. Therefore, especially for countries with poor infrastructure, investing in improvements in infrastructure may be important for attracting FDI. Nonetheless, some countries with poor infrastructure may be unattractive hosts for FDI for a variety of other reasons, and even substantial investments in infrastructure might not bring FDI pouring in. But all else being equal, a country with better infrastructure would be expected to attract more FDI (as well as more domestic investment).

The positive effect of infrastructure on FDI has been found to be quite robust to time periods and countries considered, other control variables included, the measure of infrastructure used, and the like. Examining the determinants of FDI into U.S. states for 1981 83, Coughlin, Terza, and Arromdee (1991) find that more extensive transportation infrastructures were associated with increased FDI.

Wheeler and Mody (1992) find that infrastructure quality is an important variable for developing countries seeking to attract FDI from the United States, but is less important for developed countries that already have high-quality infrastructures.

Using a self-reinforcing model of FDI, Cheng and Kwan (2000) find support for good infrastructure (density of roads) as a determinant of FDI into 29 Chinese regions from 1985 to 1995. The quality of the roads, however, did not seem to matter much: high-grade paved roads did not perform any better than all roads in determining which regions hosted the most FDI.

Infrastructure Broadly Defined Fung et al. (2005) examine whether hard infrastructure, in the form of more highways and railroads, or soft infrastructure, in the form of more transparent institutions and deeper reforms, leads to more FDI. Their analysis controls for other determinants of FDI such as regional market sizes, human capital, and tax policies. Their data are on FDI from the United States, Japan, Korea, Hong Kong, and Taiwan to regions of China. They find that soft infrastructure is a more important determinant of FDI than hard infrastructure.

Government infrastructure is used to refer to a country's political, institutional, and legal environment. It captures aspects of legislation, regulation, and legal systems that condition freedom of transacting, security of property rights, and transparency of government and legal processes (Globerman and Shapiro 2003). Government infrastructure is an important determinant of both inflows and outflows of FDI. Not only does government infrastructure attract FDI, but the proper conditions can also stimulate the creation of home-grown multinational enterprises (MNEs) that invest abroad. The biggest gains from improving government infrastructure appear to arise for small developing countries the benefits of further enhancements may be less for countries already enjoying good governance.

Globerman and Shapiro (2003) examine the effect of government infrastructure on both the probability that a country receives FDI and on the amount of FDI received (for countries receiving any FDI). They find that countries failing to achieve a mini-

mum threshold of effective governance are unlikely to receive any U.S. FDI. Thus ineffective governments that fail to promote transparent markets and whose legal systems are not rooted in English law are apt to be excluded from FDI. Globerman and Shapiro's second analysis examines the determinants of the amount of FDI, for those countries receiving FDI. They find government infrastructure, including aspects of the legal system, to be an important determinant of the amount of FDI received, for countries that do receive FDI.

As richer data become available, the influence of more types of institutions can be examined and for a wider range of countries. Given the likelihood that the impact of institutions on FDI drops off once a certain level has been achieved, most interest has focused on examining the role of institutions on FDI into developing countries. Most estimates suggest that institutions are very important. Improving institutional quality from a low level to a high level could have as much of an impact on FDI as if suddenly the country shared a border with the investing country a big change. It may be important to properly control for the correlation between per capita gross domestic product and FDI and for potential endogeneity of institutions (that FDI creates pressure for better institutions). In addition to institutions affecting FDI, some researchers have argued that FDI affects institutions, as will be discussed for the case of corruption in the next section.

Corruption Corruption is the misuse of public power or authority for private gain. Corruption tends to arise when governments control access to markets, so naturally corruption can matter for FDI. Wei (2000) established corruption's deterrent effect on FDI. Using data on bilateral investment from 12 source countries to 45 host countries, Wei finds that an increase in the corruption level in the host country leads to a reduction in inward FDI. An increase in the corruption level from that of Singapore to that of Mexico is estimated to have the same effect of deterring inward FDI as raising the tax rate by 50 percent. Additionally Wei finds that

U.S. investors are no more averse to corruption than average for investors from member countries of the Organisation for Economic Co-operation and Development (OECD).

Not just the level of corruption in a host country, but the degree to which it differs from the level in the source country, may matter for FDI. Habib and Zurawicki (2002) provide support for the negative impact on FDI of both the level of corruption in the host country and the absolute difference in the corruption level between the host and the source country. MNEs from a country with high degrees of corruption may be better able to deal with high levels of corruption in a host country than firms from a country with little corruption. The former firms are experienced in dealing with corruption, whereas the latter are accustomed to transparency. Of course, corruption may take many forms, and thus experience from one country may not fully translate to another. On the other hand, firms accustomed to dealing with bribery might be able to operate well in less corrupt environments, but even they may undergo some adjustment to the different environment. Especially when there is still some corruption, it may be hard for foreign firms to learn just where bribes are needed and where rules must be followed.

While many papers operate on the notion that corruption deters FDI, the opposite can be argued as well. In fast-growing countries with substantial bureaucracies, the ability of corruption to "grease the wheel" may be more important than the amount of the bribe required. In such situations, the bribe may be considered a small price to pay for cutting through many layers of red tape and speeding up approval. When weighing costs versus benefits, how big a bribe is required must be compared to how much improvement in speed or likelihood of approval is gained. The terms "helping hand" versus "grabbing hand" corruption have been used to distinguish corruption that positively affects FDI from that which negatively affects FDI (see Egger and Winner 2006). The fact that corruption on the whole seems to deter FDI suggests that corruption does more grabbing than helping.

In addition to the question of how corruption affects FDI there is also the question of how FDI might affect corruption. Under the U.S. Foreign Corrupt Practices Act of 1977, U.S. firms are not allowed to bribe (or give gifts to) foreign governments to gain favor in contracts, so FDI from the United States might be expected to push toward less corruption. Kwok and Tadesse (2006) propose three avenues through which MNEs may affect institutions in host countries. First, the regulatory pressure effect describes implications of the fact that foreign firms may be constrained to not pay bribes. Bribes may be against company norms, rules set by the source government, or conventions set by the global business community. Second, the demonstration effect is based on the notion that, like productivity, the tendency of MNEs to avoid corruption may spill over to other firms. When local firms deal with MNEs or hire some of their former workers, they can observe how business decisions are made in MNEs. The presence of MNEs should counter existing norms by demonstrating an alternative method of conducting business that can be more efficient. Finally, the professionalization effect relates to the likelihood that leaders (or future leaders) of host firms will acquire training in professional business practices (which discourage corrupt practices) and that these new practices will become socialized in younger generations.

One should recognize that the effect of corruption may be difficult to separate from other aspects of government infrastructure such as bureaucracy, as corruption and bureaucracy tend to be linked (corruption arises to cut through the bureaucracy). In addition to affecting whether a country receives FDI and how much FDI it receives, corruption could also affect the value foreign (or even domestic) firms are willing to pay when acquiring local firms. In general, corruption may be one of many dimensions a country may seek to improve in the hope of becoming a more attractive location for FDI.

Remaining Questions More work needs to be done on how firms adapt to environments plagued by corruption. Some evidence suggests that more firms may opt for joint ventures in the face of corruption. A local partner may be more experienced at dealing with corruption and the host government in general. What other strategies do (or should) firms use when operating in a corrupt environment? Similar to joint ventures, are firms more likely to opt for acquiring a local firm over greenfield investment (building a new plant) in corrupt environments? Is there evidence that firms adopt an expansion strategy in which experience in moderately corrupt countries helps prepare them to begin operations in more severely corrupt regimes? Do firms hire a larger proportion of local workers in more corrupt environments?

More work should also be done to address the variation in types of corruption. All corruption is not created equal. Are some forms of corruption more damaging to FDI than others? Some forms of corruption may act more like fixed costs, such as a bribe for approval to enter the market. Such corruption could be less distorting (as long as it is not prohibitive) than a bribe that is set in relation to number of employees, production, or profits. What evidence is there that bigger or more profitable firms are expected to pay larger bribes? In which situations are foreign firms especially exposed to corruption, and when are domestic firms just as bad off?

Kellenberg (2007) compares a policy of public input provision to a policy of subsidy incentives for attracting FDI. More analysis of this kind is needed, such as comparing a reduction in corruption or an improvement in government infrastructure to use of subsidies. One might like to know, due to corruption (alone), how much of a subsidy would Mexico have to pay to make it as attractive to FDI as Singapore? The next decade of research should bring us answers to such questions, and many more.

See also corruption; location theory

FURTHER READING

Cheng, Leonard K., and Yum K. Kwan. 2000. "What Are the Determinants of the Location of Foreign Direct Investment? The Chinese Experience." *Journal of International Economics* 51 (2): 379 400. Evidence that good

infrastructure, measured as density of all roads, matters for FDI into Chinese regions.

Coughlin, Cletus C., Joseph V. Terza, and Vachira Ar romdee. 1991. "State Characteristics and the Location of Foreign Direct Investment within the United States." *Review of Economics and Statistics* 73 (4): 675 83. Finds that more extensive transportation infrastructures were associated with increased FDI into U.S. states.

Egger, Peter, and Hannes Winner. 2006. "How Corrup tion Influences Foreign Direct Investment: A Panel Data Approach." *Economic Development and Cultural Change* 54 (2): 459 86. Argues that corruption deters FDI in developing but not developed countries and that the importance of corruption has declined over time.

Fung, K. C., Alicia Garcia Herrero, Hitomi Iizaka, and Alan Siu. 2005. "Hard or Soft? Institutional Reforms and Infrastructure Spending as Determinants of Foreign Direct Investment." *Japanese Economic Review* 56 (4): 408 16. Demonstrates that soft infrastructure can be more important than hard infrastructure for attracting FDI to Chinese regions.

Globerman, Steven, and Daniel Shapiro. 2003. "Govern ance Infrastructure and U.S. Foreign Direct Invest ment." *Journal of International Business Studies* 34 (1): 19 39. Finds government infrastructure matters for whether a country receives any FDI and also for the amount of FDI received.

Habib, Mohsin, and Leon Zurawicki. 2002. "Corruption and Foreign Direct Investment." *Journal of International Business Studies* 33 (2): 291 307. Considers both the level of corruption and the absolute difference in the levels across countries.

Kellenberg, Derek K. 2007. "The Provision of Public Inputs and Foreign Direct Investment." *Contemporary Eco nomic Policy* 25 (2): 170 84. Compares investment in public infrastructure to subsidies as methods for at tracting FDI.

Kwok, Chuck C. Y., and Solomon Tadesse. 2006. "The MNC as an Agent of Change for Host country In stitutions: FDI and Corruption." *Journal of Interna tional Business Studies* 37 (6): 767 85. Considers three avenues through which FDI can deter corruption.

Wei, Shang Jin. 2000. "How Taxing Is Corruption on In ternational Investors?" *Review of Economics and Statistics* 82 (1): 1 11. Compares the deterrent effect of corrup tion to that of higher taxes, and compares the response of FDI from the United States to corruption to that typical for the OECD.

Wheeler, David, and Ashoka Mody 1992. "International Investment Location Decisions: The Case of U.S. Firms." *Journal of International Economics* 33 (1 2): 57 76. Finds that the quality of infrastructure is quite im portant for U.S. firms locating in developing countries, but not for developed country hosts.

AMY JOCELYN GLASS

■ intangible assets

The integration of the world economy has under lined the need for firms to exploit their intangible assets on a global scale. Firms' intangible assets in clude their stock of knowledge, which is related to ideas, research and development (R&D), patent and blueprints, scientific and technical workers, and management techniques. In addition, intangible as sets consist of their stock of goodwill, which is asso ciated with product quality reputation, trademarks, and brand names.

The empirical evidence shows that the growth in intangible assets has been a recent phenomenon. As emphasized by Griliches (1994), the source of eco nomic growth and wealth lies no longer in the in vestment of physical, tangible assets but in the crea tion and use of intangible assets. The contribution of the latter is hard to measure due to lack of data and uniform definition concerning its measurement. There is, however, substantial anecdotal evidence that enterprises have increased their relative invest ment in intangible assets. The pace of intangible investment by U.S. private firms has risen sharply in recent decades. By the late 1990s, their investments in intangible assets were around 1 trillion dollars a year, which was about the same as expenditure on tangible assets (Nakamura 2001).

The traditional explanation of multinational en terprises (MNEs) rests on the existence of firm-spe cific intangible assets. These assets give MNEs off setting advantages over foreign producers because

they have the property of *joint inputs*. Relative to physical tangible assets, intangible assets are easier and cheaper to transfer to foreign subsidiaries. For instance, blueprints and patents are very costly to produce but once they are created, they can be supplied to any foreign subsidiaries without reducing their value or productivity. Investing in tangible assets typically involves fixed costs at the plant level, whereas investing in intangible assets involves fixed costs at the level of the firm.

The empirical evidence confirms that MNEs have a larger value of intangible assets relative to their market value (compared to other firms). Industries with a high proportion of multinational firms tend to be characterized by high levels of intangible assets (evidenced as substantial R&D and advertising expenditures relative to sales), new or sophisticated products, and high shares of skilled employment (Markusen 1995). Since the early 1990s, for instance, the stock of business services has risen ninefold, to reach 26 percent of total inward foreign direct investment (FDI) stock in services in 2002 (UNCTAD 2004).

Understanding the importance of intangible assets for FDI requires investigating their transferability to local firms and their impact on host economies.

Transferability of Knowledge-Based Assets Transferring knowledge-based assets to local firms is quite different from transferring other intermediate inputs. Knowledge is not necessarily easily transferred, especially across firms. The reason is that knowledge-based assets are usually deeply embedded in the institutional and organizational contexts of their origins. Moreover, knowledge-based assets have a tacit component that is not codified and therefore is imperfectly transferable to local firms. For instance, production processes requiring highly specialized workers are usually transferred to subsidiaries rather than to other firms since licensing would entail costly training of local employees (Teece 1986).

The transmission of knowledge-based assets is subject to market failures such as the dilution of property rights. Product reengineering, for example, may allow the local firm to discover the tacit knowledge embedded in the new product and to start

a rival firm. In addition, the MNE might break the terms of the license and transfer the knowledge-based asset to a different local firm. The problem for the MNE is to design an optimal contract that preserves the value of its knowledge-based assets. This contract usually includes rent sharing with the local firm. If defection cannot be avoided, the MNE will prefer to internalize its assets in spite of higher setup costs.

The empirical evidence presented by Smith (2001) on U.S. firms' foreign operations shows that licensing is a more likely entry mode in markets characterized by stronger protection of intellectual property rights.

Transferability of Goodwill Assets The stock of knowledge is not the only intangible asset that is difficult to transfer to local firms. The transfer of goodwill assets such as reputation for quality can face severe problems if the local firm does not fully appropriate the returns from maintaining the MNE's brand reputation. Examples of industries where reputation is important include hotels and restaurants, consulting, and financial services. The reputation problem faced by MNEs is as follows. Consumers value quality and are willing to pay more if they believe that the firm is supplying high-quality products or services. The local firm may be tempted to compromise on products' quality to reduce costs. This strategy leads to higher short-term profit, but once consumers become disappointed with their purchases, the MNE may retaliate and purchase products from other suppliers.

The fact that foreign parties do not fully appropriate the returns from maintaining the MNE's reputation gives rise to FDI. The local firm does not internalize the costs to the global reputation of the MNE when it provides low-quality goods. This problem is why the MNE should design contracts that specify much more than prices. In practice, the usual method to achieve such a result is a franchise contract in which the local firm agrees on a complete business concept. The contract needs to ensure that the local firm provides a level of quality sufficient to maintain the reputation of the MNE. However, the franchising agreement will still not be sufficient to

solve the reputation problem when efforts to maintain the quality of products and/or services are not fully verifiable. In this case, the MNE may prefer to set up its own subsidiary, which allows more effective incentive and control systems.

Intangible Assets and Spillovers to Foreign Economies Researchers have attempted to identify and empirically measure spillovers from MNEs' transfer of intangible assets to local firms. These spillovers may be positive or negative. On the one hand, MNEs may enable local firms to upgrade their technology, to the extent that they bring better business practices, technology, or management. On the other hand, MNEs may increase competition by attracting demand away from local firms. If economies of scale are important, the loss of demand will reduce the productivity of local firms by reducing the scale of their production.

Positive or negative spillovers may be direct, from firm to firm, through imitation of technology, managerial and organizational innovations, or competition. They can also be indirect through the labor market when specialized skilled workers from the MNE are complementary to workers in the local firm or when they move to rival firms. In either case, the ability of local firms to assimilate or value new knowledge, the so-called *absorptive capacity*, is crucial for obtaining significant benefits from FDI.

There is weak evidence for spillovers from MNEs. Positive spillovers from multinational firms are supported by casual evidence from many countries, but their existence and magnitude are difficult to establish empirically (UNCTAD 2001). Several studies have emphasized the role of local firms' absorptive capacity in explaining the lack of spillovers from multinational activities. This reasoning suggests that spillovers may not affect firms equally but may benefit only firms with high levels of absorptive capacity. The lack of strong empirical evidence for spillovers suggests also that multinational firms have succeeded in protecting their intangible assets.

In sum, intangible assets are vital to the incentives for MNEs to form and vital to the ability of MNEs to continue to prosper. Successful MNEs have valuable intangible assets, transfer the benefits of these intangibles to subsidiaries producing abroad, and protect their intangible assets by limiting the degree that the benefits spill over to rival firms.

See also foreign direct investment: the OLI framework; technology spillovers

FURTHER READING

Griliches, Zvi. 1994. "Productivity, R&D, and the Data Constraint." *American Economic Review* 84 (1): 1 23. An empirical analysis that questions the determinants of growth in light of the increase of investment in intangible assets such as R&D.

Markusen, James. 1995. "The Boundaries of Multinational Enterprises and the Theory of International Trade." *Journal of Economic Perspectives* 9 (2): 169 89. Survey of the ownership, location, and internalization advantage of MNEs that incorporates new theories of international trade.

Nakamura, Leonard. 2001. "What Is the U.S. Gross Investment in Intangibles? (At Least) One Trillion Dollars a Year!" Working Paper No. 01 15. Federal Reserve Bank of Philadelphia. Suggest an estimation of U.S. investment in intangible assets.

Smith, Pamela. 2001. "How Do Foreign Patent Rights Affect U.S. Exports, Affiliate Sales, and Licenses?" *Journal of International Economics* 55 (2): 411 39. An empirical study finding that stronger intellectual property rights protection leads to more licensing.

Teece, David. 1986. *The Multinational Corporation and the Resource Cost of International Technology Transfer.* Cambridge, MA: Ballinger. Analysis of knowledge transfer costs, market externalities, and the internalization of knowledge assets.

United Nations Conference on Trade and Development (UNCTAD). 2001. "Should Countries Promote Foreign Direct Investment?" G24 Discussion Paper Series No. 9. Geneva: UNCTAD. Determinants and impacts of spillovers from FDI and their policies implications.

. 2004. "The Shift toward Services." *World Investment Report.* Geneva. UNCTAD. Available at http://www.unctad.org/en/docs/wir2004 en.pdf. Annual overview of trends in FDI, including a detailed analysis of FDI in intangible assets.

FARID TOUBAL

■ **integrated framework**
See trade-related capacity building

■ **intellectual property rights**
Intellectual property rights are legal devices designed to protect intellectual property, which can be broadly defined as a recognized ownership over the ideas, designs, inventions, or concepts created by a person or organization. Intellectual property rights are therefore instruments intended to ensure that the originators of intellectual property have control over its use.

Intellectual property rights relate to two broad categories of intellectual property: industrial property, and literary and artistic property. Industrial property can be protected by patents, utility models, industrial designs, trademarks, and geographical indications. Literary and artistic property protection may be provided by copyrights. Intellectual property rights may also be granted in the form of sui generis protection. Additionally, trade secrets offer an alternative way of providing protection to intellectual property.

Types of Intellectual Property Rights *Patents* allow an inventor to exclude all others from commercially exploiting his or her invention for a fixed period of time, usually 20 years. They apply both to novel products (such as new drugs) and to novel processes (such as a new method of producing an existing drug).

Utility models are also used to protect inventions in certain countries but are usually sought for less technically complex devices. The requirements for achieving protection under utility models are less stringent than the ones for achieving protection under patents, and the length of protection is also shorter for utility models.

Industrial designs protect the ornamental or aesthetic features (shape, pattern, color) of a useful article that is reproduced by industrial means.

Trademarks and service marks are used to protect a distinctive mark or name that identifies a product, service, or company.

Geographical indications are signs used on goods that originate from a particular geographical location and possess qualities or a reputation that are due to that location.

Copyrights protect the rights of creators of literary and artistic works (such as books, poems, music, paintings and sculptures, films) to communicate, display, perform, and make or sell copies of those works. The length of protection is usually the lifetime of the creator plus 50 to 70 years. Copyrights may also apply to technology-based works such as computer programs, electronic databases, and multimedia productions.

Sui generis protection refers to new categories of intellectual property rights that apply to forms of creative activities resulting from recent technology advances that do not fit into the standard categories of intellectual property. Computer software, layout design of integrated circuits, plant breeder's rights, and electronic transmission of databases or broadcasts are some examples.

Trade secrets are not intellectual property rights in the traditional sense. They provide protection to industrial property through legislation that deems it illegal to disclose or use proprietary information about production practices and processes that is considered commercially valuable for a company (ranging from a list of secret ingredients to customer lists). Trade secrets have no statutory time limit but the protection they offer may expire through learning by fair means, such as reverse engineering or reading public documents.

National and International Dimensions of Intellectual Property Rights Intellectual property rights are national in scope: legislation regarding intellectual property rights is enacted at the national level and is part of a country's legal system. Countries face a fundamental trade-off when designing a system of intellectual property rights. On the one hand, intellectual property rights should generate a sufficient incentive for invention and creation to occur by providing an effective means of protection to the rights of the originators of intellectual property. On the other hand, intellectual property rights should ensure proper diffusion and access to intellectual

property, so that society as a whole benefits from public access to new ideas. For instance, by offering intellectual property protection through a patent, society has opted to create a temporary monopoly (or exclusive market) for the product protected under that patent. The presence of such monopoly imposes societal costs in terms of high prices and lower efficiency, but it also generates the incentives for invention to occur by guaranteeing higher profits to patent holders. Once a patent is granted, however, all technical information pertaining to that new product (or process) becomes publicly available for others to use in their own research activities. In this sense, intellectual property rights restrict the use of proprietary ideas in production, but patenting may foster their use by others in research. Therefore, the societal benefits derived from having access to this information may outweigh the costs imposed on society by the presence of a monopoly.

Economists explain this basic trade-off by thinking about intellectual property rights in terms of the provision of a public good. First, intellectual property is nonrivalrous, that is, one person's use of it does not reduce another person's use. Therefore the marginal cost of providing access to the information embodied in intellectual property to another user is very small and has the potential to generate benefits to society as a whole. For example, the patenting of a plasma display panel can generate positive benefits (or externalities) through technological spillovers, since all the technical details pertaining to such invention can now be accessed by other researchers at a negligible cost. Second, intellectual property may be nonexcludable by private means. That is, it may be impossible to preclude others from using the information without authorization. This generates what economists refer to as a free-rider problem: without protection of intellectual property, no one will wish to bear the costs of generating it.

Most legal systems try to balance these two goals of providing financial incentives for innovation and providing society with access to new ideas and inventions, which is a challenging task. If intellectual property rights were too weak, plasma display

panels might have never been created. If intellectual property rights were too strong, the spillover benefits associated with such an invention might never sufficiently spread across society. Note that this basic trade-off presents static and dynamic considerations. By bearing the static costs of protecting intellectual property today, society may enjoy the dynamic gains of faster innovation and growth in the future.

Each form of intellectual property right is designed with these goals in mind. When patents are granted, society gains full access to the technical information regarding a new product or process, and the fixed length of protection minimizes the costs of providing access to this product or process at prices higher than marginal costs. Copyright protection allows exceptions under the "fair use doctrine," which defines activities that can make use of protected works in the interests of educational, scientific, and technical advances. It is less clear how trade secrets can provide net benefits to society since they generate market power without divulging information. Economists have started to recognize, however, that protecting trade secrets may also generate important incentives for innovation, for example, by reducing incentives for research and development patent races or by facilitating reverse engineering.

Additional considerations about designing a system of intellectual property protection are relevant to the case of an open economy. For countries that are net exporters of intellectual property, large losses can occur from the imitation and use of their innovations in foreign countries with weak intellectual property rights. For countries that are net importers of products subject to intellectual property protection, the decision to provide or even strengthen intellectual property rights generates static costs such as higher prices for imported goods or technology, transfer of monopoly rents to foreign firms or foreign countries, and loss of employment in local firms whose production is imitative in nature. And to the extent that imitative activities may generate technological spillovers and affect a country's growth rate, intellectual property protection may impose dynamic costs in

terms of reduced access to international technologies and slower growth in the long run.

Stronger intellectual property rights may also result in dynamic welfare gains. For instance, the transfer of monopoly rents to foreign firms may result in innovation that is better suited to domestic needs than to foreign needs. Additionally, stronger intellectual property rights can facilitate technological diffusion, innovation, and growth by generating incentives for trade in knowledge-intensive goods, foreign direct investment (FDI), and licensing agreements. A key issue in evaluating the net impact of these trade-offs for an open economy is a country's overall technical potential. If the local economy does not possess the infrastructure and know-how to reproduce new technologies through imitation or reverse engineering, weaker intellectual property protection would not generate the positive effects of technological diffusion. In this case, stronger intellectual property rights might be more desirable since they provide better opportunities for formal technology transfer through access to knowledge-intensive goods, foreign patents, joint ventures, and licensing agreements. Another important consideration is market size; foreign firms may not have enough incentives to engage in production targeted to domestic needs if the size of the local market is relatively small.

Since intellectual property rights are national in scope, the level of protection provided by such instruments (and the level of enforcement of such protection) can vary widely across countries, with the largest variations occurring between developed and developing countries. These discrepancies exist not only with respect to the de jure level of protection, but also with respect to the de facto enforcement of the existing legislation. This issue becomes especially important and controversial as owners of intellectual property seek to exploit international markets. The debate is relevant not only for individuals or organizations that possess ownership of intellectual property but also for national policy, since a country's level of intellectual property rights may affect its volumes of international trade and FDI flows and its levels of innovation and economic growth.

The International Debate Regarding Intellectual Property Rights The need to protect industrial property at the international level was first recognized in the Paris Convention for the Protection of Industrial Property in 1883, and later extended to literary and artistic property in the Berne Convention for the Protection of Literary and Artistic Works in 1886. These treaties, which constitute the core foundation of the World Intellectual Property Organization (WIPO), established a system in which inventors or artists could seek protection for their work in other countries on terms no less favorable than the terms available to the nationals of those countries. Note that this would imply that a U.S. patent holder, seeking to extend protection of her intellectual property to other countries, would still need to apply for a patent in each country separately. The treaties, however, guaranteed that she would be treated just like any other citizen of that country. Since then, the international legislation regarding intellectual property rights has undergone several changes primarily driven by two important forces: technological progress and globalization. First, technological progress has required the adaptation of intellectual property rights instruments in key areas such as integrated circuits, computer software, and biotechnology inventions. Second, the globalization process has facilitated the diffusion of intellectual property across international borders, increasing the controversy over discrepancies in intellectual property rights across countries.

The current system of global intellectual property rights encompasses various intergovernmental treaties administered by the WIPO, as well as intellectual property rights legislation that is part of bilateral, regional, or international trade agreements. The most important and controversial of such agreements is the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), which came into effect in 1995 under the auspices of the World Trade Organization. In contrast to the Paris Convention, the TRIPS agreement imposes minimum standards of protection, which are either in line with or exceed the practices of most industrialized countries, and

establishes a dispute settlement mechanism to enforce these minimum standards.

As intellectual property rights become a standard feature in many international trade agreements, the need for a better understanding of how these rules will affect countries becomes a pressing issue in economic research. Recent theoretical and empirical research has focused on trying to better understand the specific channels through which intellectual property rights might affect innovation and economic growth, such as international trade flows, FDI, and licensing.

Economic theory suggests an ambiguous relationship between intellectual property rights and international trade flows. On the one hand, stronger intellectual property rights raises demand for the protected goods and could lead to increased sales. On the other hand, stronger intellectual property protection also augments the market power of foreign patent holders, allowing them to charge higher prices, which could generate reduced sales. Empirical studies in this area are therefore needed to settle this question.

Research undertaken so far suggests that stronger intellectual property rights have a positive impact on overall trade volumes, but are not relevant to trade in high-technology goods. Several explanations have been offered for this puzzle. It is possible that high-technology trade flows are less responsive to intellectual property rights because those products are simply harder to imitate than low-tech products. It is also conceivable that intellectual property rights matter the most to middle-income large developing countries, which are more likely to possess the technical capabilities to undertake reverse engineering. Finally, high-technology firms may decide to serve foreign markets through FDI and licensing.

The theoretical ambiguity with respect to the impact of intellectual property rights also extends to FDI and licensing. Empirical studies suggest that intellectual property rights are less important in attracting FDI than other factors, such as a country's infrastructure or overall business climate. Stronger intellectual property rights, however, are important in stimulating formal transfers of technology, espe-cially in the case of large developing countries with strong abilities to absorb technology. This result also holds for international licensing flows, since stronger intellectual property rights reduce the costs of establishing and monitoring licensing contracts.

Economic research in all these areas relies on national rankings of intellectual property rights, mainly the Ginarte and Park (1997) index. This is a time-varying index of intellectual property rights protection based on five categories of patent law: extent of coverage, membership in international patent agreements, provisions for loss of protection, de jure enforcement mechanisms, and the duration of protection. The index provides extensive coverage in terms of countries and years; however, one of its main drawbacks is the fact that it does not account for the de facto enforcement of the legislation.

Other Controversial Issues The debate over intellectual property rights is further complicated by ongoing deliberations regarding such critical issues as parallel imports, compulsory licensing, and traditional knowledge.

Parallel Imports Parallel imports occur when price differences between countries, generated by disparities in intellectual property protection, allow international arbitrage to take place. Consider a situation in which legitimate goods (as opposed to knock-offs) are placed in circulation in a certain market under legislation that protects intellectual property. If the wholesale price of such goods is low enough and transportation costs are not prohibitive, it may be possible for an independent trader to profit by selling those goods in other markets in which intellectual property rights are weaker. Parallel imports are not prohibited under international legislation (TRIPS), and their legality depends on whether or not national laws confine protection to the country in which the protected goods are first sold. If protection is exhausted at the international level, firms fully lose control over a product's distribution once the product has been put for sale in any location. In that case, markets are open to parallel imports from foreign countries. Economists have shown that the effects of limiting parallel imports on a country's

welfare are theoretically ambiguous and depend on the underlying motivations to pursue parallel trading in the first place.

Compulsory Licensing Compulsory licensing refers to a situation in which the government of the importing country has the legal authority to force the foreign holder of a patent to license production to a local firm as a condition of patent protection. This practice is allowed by international legislation (TRIPS) only under extreme circumstances, such as public health emergencies.

The regulation of parallel imports and the use of compulsory licensing have been critical issues in various international forums on providing access to medicines in developing countries.

Traditional Knowledge Most research in economics has focused on patents and copyrights, ignoring questions related to traditional knowledge. Traditional knowledge generally refers to the wisdom, knowledge, and teachings that are long-standing traditions of certain regional, indigenous, or local communities. These traditions can be expressed through various means, such as stories, folklore, rituals, and songs. The controversy over the protection of traditional knowledge has its roots in the 1992 Convention on Biological Diversity, which recognized the value of traditional knowledge in protecting species, ecosystems, and landscapes in language that contradicts international intellectual property agreements, especially the TRIPS agreement. Protecting traditional knowledge requires a sui generis type of protection and is a particularly important issue in developing countries, which have objected, for example, to patenting of traditional uses of medicinal plants.

Future Concerns The scope and pace of globalization, as well as technological progress, will ensure that intellectual property rights continue to play an important role in negotiations regarding international trade agreements. Intellectual property rights will also remain an important issue in designing domestic policies, especially in developing countries, since they have the potential to affect important macroeconomic variables such as a country's growth rate, volume of international trade, or FDI and licensing flows. Economists' efforts to better understand the precise mechanisms through which intellectual property rights affect these variables are therefore essential in facilitating the resolution of international disputes over intellectual property protection. But this research effort should also try to improve existing measures of cross-country comparisons involving intellectual property rights, and focus on less explored issues, such as whether and how to provide intellectual property protection to traditional knowledge.

See also access to medicines; Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS); Convention on Biological Diversity; foreign direct investment (FDI); intellectual property rights and foreign direct investment; World Intellectual Property Organization; World Trade Organization

FURTHER READING

Finger, J. Michael, and Philip Schuler. 2003. *Poor People's Knowledge: Promoting Intellectual Properties in Developing Countries.* Washington, DC: World Bank and Oxford University Press. Addresses questions of traditional knowledge, access to genetic resources, and other topics that are increasingly relevant to policymakers in developing countries.

Fink, Carsten, and Keith Maskus, eds. 2005. *Intellectual Property and Development: Lessons from Recent Economic Research.* Washington, DC: World Bank and Oxford University Press. This book provides a compilation of articles on various controversial issues related to intellectual property rights.

Ginarte, Juan Carlos, and Walter G. Park. 1997. "Determinants of Patent Rights: A Cross National Study." *Research Policy* 26 (3): 283 301. Provides details on the criteria used to construct the Ginarte Park index.

Maskus, Keith. 2000. *Intellectual Property Rights in the Global Economy.* Washington, DC: Institute for International Economics. Even though the main focus of this book is the TRIPS agreement, chapters 3, 4, and 5 offer a broad overview of the economics of intellectual property rights and their impact on the world economy.

Owen, Lippert. 1999. *Competitive Strategies for the Protection of Intellectual Property.* Vancouver, Canada: Fraser

Institute. Provides a compilation of articles that relate to the many facets of designing a system of intellectual property rights.

PATRICIA HIGINO SCHNEIDER

■ intellectual property rights and foreign direct investment

The impact of intellectual property rights (IPRs) on foreign direct investment (FDI) plays a central role in the debate between developed and developing countries over intellectual property rights protection. Many developing countries have begun to reform their IPR regimes in response to the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) and pressure from developed countries. As multinational enterprises (MNEs) strive to make use of their knowledge assets beyond national borders, they have made IPRs an issue of multilateral negotiations. Governments of developing countries also believe that agreeing to TRIPS ultimately would allow them wider access to agricultural and textile markets in developed countries. In addition, business interests within many developing countries have encouraged their governments to adopt stronger IPR protection in order to protect their own innovations tailored to the domestic market. Some observers also argue that stronger IPRs protection can encourage imports, inward FDI, and technology licensing, all of which can lead to an increase in technology transfer.

Effect of IPRs on Technology Transfer and Innovation Technology is fundamental to economic development, and its importance for raising productivity and living standards has long been recognized. Innovation and technological progress can raise productivity through the introduction of new goods, the improvement of existing goods, and reduction in the costs of production. Resources for innovations tend to be highly concentrated in a small number of advanced countries, however. For developing countries whose firms are not at the technological frontier, international technology transfer is important for promoting development. Some transfers occur between willing partners in formal trans-actions, but much comes through informal non-market transactions or spillovers. One of the transfer channels is trade in goods and services, with imports of goods having the potential to transfer knowledge through reverse engineering and cross-border learning of production methods and product design. Another channel is FDI, particularly inward FDI, where MNEs deploy advanced technology to their subsidiaries that may be diffused to host-country firms. Licensing, which involves the purchase of production and distribution rights for a product and the knowledge required to make effective use of these rights, is another channel for technology diffusion. Joint ventures combine many of the properties of FDI and licensing, and hence will also involve technology transfer. Informal channels of technology diffusion include imitation, movement of personnel from one firm to another taking with them specific knowledge of their original firm's technologies, data in patent applications, and the migration of people. There is no formal compensation to the original owner of the technology transferred. In this case, IPRs clearly play a role.

Intellectual properties have the characteristic of a public good: they are available on a nonexclusive basis. It is not possible to prevent others from applying knowledge even without the authorization of its creator. They are also nonrivalrous in use (additional parties can benefit from them at zero additional cost). Once an innovation has been created, its nonrival character suggests that benefits will be maximized if its use is free to all at marginal cost. Although a policy of free access might yield benefits in the short run, it will damage the incentive for further innovation. When the new technology is imitated, it reduces the potential profits of the original inventor and potentially removes the incentive to engage in innovative activities. Since imitation has lower costs than innovation, imitators have the advantage over innovators unless the latter can restrict access to their innovation. This characteristic provides the argument for strong IPRs protection. In the absence of IPRs, the market for intellectual properties would fail or yield an inefficient supply of output. As to the connection to FDI and trade, many pro-

ducers of intellectual output are engaged in both domestic and foreign markets. Risks of unauthorized copying and imitation exist both at home and abroad. Thus in regions where IPRs are weak, incentives to market via trade or FDI might be weak since the innovating firm's ability to appropriate rents is threatened. Moreover, an infringement of IPRs may adversely affect incentives to innovate and produce and thereby affect the potential to export and invest abroad.

Effect of IPRs on Imports, FDI, and Licensing
In determining the impact of IPRs on FDI, it is useful to rely on the widely accepted ownership-location-internalization (OLI) framework by Dunning (1993). When selling its products abroad, a firm is at least initially disadvantaged relative to local producers. Thus, to compete effectively with indigenous firms, a foreign producer must possess some advantages. Ownership advantage explains that MNEs must have assets that are unique to the firms, which usually take the form of new technologies, know-how, reputation for quality, or brand name. Ownership advantages assure a firm's ability to enter the host country's market. Location advantages are needed for firms to locate business abroad. This could be in the form of lower transportation cost, avoiding tariffs, lower input prices, lower standards and regulations, and access to the distribution markets. Internalization advantage gives the incentive to the MNEs to retain full control over the production process rather than to license their knowledge assets to local firms in a foreign country. This advantage covers the costs from license contract breaching, as well as control over the quality and reputation of the MNEs' products.

As mentioned earlier, MNEs have several options to exploit their knowledge assets. Apart from FDI, international trade and cross-border licensing represent the most important channels to transfer technology to the host countries. Hence, according to the OLI framework, IPRs protection should have an impact on the relative attractiveness of the three modes. Weak IPRs protection increases the probability of imitation, which erodes a firm's ownership advantages and makes a host country a less attractive location for foreign investors. An inadequate IPRs regime, therefore, deters FDI and encourages exporting. At the same time, a weak IPRs system increases the benefits of internalization since it is associated with a greater risk of the licensee's breaching the contract and acting in direct competition with the seller; this improves the attractiveness of FDI while discouraging licensing. Thus the overall relationship between the level of IPRs protection and FDI can be ambiguous.

Other theories conclude that the impact of IPRs protection on FDI varies by industries. MNEs are more likely to undertake FDI rather than licensing when they have a complex technology and highly differentiated products and when the costs of transferring technology through licensing are high (Horstmann and Markusen 1987). In such circumstances stronger IPRs, by reducing the risks of technology leakage, may increase the extent of licensing (Yang and Maskus 2001), thus reducing the need for FDI. For industries where technologies are relatively easy to imitate, however, we would expect more attention to be paid to the strength of IPRs protection. As such, the IPRs regime is likely to be important for sectors such as drugs, cosmetics, and health care products; chemicals; machinery, and equipment; and electrical equipment (Smarzynska 2004).

The concern about the IPRs regime also depends on the purpose of an investment project, being the highest in the case of R&D facilities and the lowest for projects focusing exclusively on sales and distribution (Mansfield 1994). This difference is expected given the nature and function of IPRs, which center on the ability to appropriate. In industries where firms have alternative means to appropriate the returns to their investments, the firms may forgo seeking IPRs and rely on other forms of protection. Other methods of protecting their knowledge assets include: (1) education and incubation of local innovators: for example, Microsoft is helping China develop its local software industry; (2) R&D strategies: firms increase their R&D to increase their lead time, sales, and services because litigation is slow, costly, and uncertain (some of them add anticounterfeiting features to prevent imitation); (3) other

business strategies: working with the imitators, lower prices on products in developing countries, and bundling products with services. These possibilities might explain why FDI in certain industries is found to be insignificantly affected by IPRs.

Empirical evidence on the relationship between IPRs protection and FDI is mixed. Lee and Mansfield (1996) find positive relationship between the two in 14 developing host countries, while Primo Braga and Fink (1998) do not obtain statistically significant results. When considering the simultaneous effects of IPRs on three entry modes (export, FDI, and licensing), Smith (2001) found that strong IPRs promote licensing more than FDI but have no impact on exports, while Maskus, Saggi, and Puttitanun (2005) found that strong IPRs result in a higher probability of FDI relative to licensing.

The inconsistency of empirical results might be due to the crude measure of IPRs protection. The measurement problems are common in proxy for the IPRs protection such as national patent indexes of Rapp and Rozek (1990) or Ginarte and Park (1997). While it might be relatively easy to classify relevant laws and regulations on the books (such as coverage and patent length), actual enforcement is very difficult to judge objectively. Moreover these measures capture only overall countrywide variations in the protection of intellectual property in a single national index, ignoring the cross-industry differences in effective protection.

Conflicts between Developed and Developing Countries over IPRs Strengthening the IPRs regime has been pushed aggressively by the United States and other major industrialized countries. These developed countries tend to opt for relatively strong IPRs systems, with the aim of encouraging creative activities and innovations that are seen as an important source of long-run economic growth.

Developing countries, on the other hand, argue that stronger IPRs could limit the local imitation of new technologies, reduce the supplies of products, and increase the prices of new goods. Hence the majority of developing countries have opted for weak IPRs protection, if any, as a way of allowing the rapid diffusion of knowledge through imitation as a sig-

nificant source of technological development. Providing stronger IPRs protection is seen as shifting profits from domestic imitative firms to foreign firms and reducing outputs of new technologies to the monopoly level, and hence increasing prices in the domestic economy rather than encouraging domestic innovative activity.

The counterargument is that stronger IPRs protection can help reward creativity and risk-taking even in developing economies, while weak IPRs protection can keep developing countries dependent on dynamically inefficient firms that rely on counterfeiting and imitation. It is also argued that the lack of technology transfers to the developing countries via FDI is due to the inadequacy of the IPRs protection and that such protection would promote FDI.

Nevertheless, economics studies showed that the net welfare effect of the host country depends on the host country's characteristics. In a small country where production and innovation capabilities are limited, higher IPRs standard likely will improve welfare through higher access to products that would not be available otherwise. If a country has greater production and imitation capability, but limited innovative capacity, higher standards of protection will hurt local producers and result in lower welfare for the economy. However, as innovative capacity improves, a country will gain from higher IPRs standard (Chen and Puttitanun 2006). In sum, strengthening IPRs protection has different welfare implication depending on the characteristics of each country.

See also Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS); foreign direct investment (FDI); foreign direct investment and innovation, imitation; foreign direct investment and international technology transfer; foreign market entry; fragmentation; internalization theory; international investment agreements; location theory; multinational enterprises; technology licensing; technology spillovers; trade-related investment measures (TRIMs)

FURTHER READING

Chen, Yongmin, and Thitima Puttitanun. 2005. "Intellectual Property Rights and Innovation in Developing

Countries." *Journal of Development Economics* 78: 474 93. The first formalized model showing a U shaped re lationship between development level and IPRs.

Dunning, John H. 1993. *Multinational Enterprises and the Global Economy.* Suffolk, UK: Addison Wesley. This seminal work explains advantages needed to become a successful MNE.

Ginarte, Juan C., and Walter Park. 1997. "Determinants of Patent Rights: A Cross National Study." *Research Policy* 26: 283 301. Developed a widely used IPRs index that expanded on the work by Rapp and Rozek.

Horstmann, Ignatius, and James R. Markusen. 1987. "Li censing versus Direct Investment: A Model of Inter nalization by the Multinational Enterprise." *Canadian Journal of Economics* 20: 464 81. First paper to formalize the notion of internalization incentives for multinational firms.

Lee, Jeoung Yeon, and Edwin Mansfield. 1996. "In tellectual Property Protection and U.S. Foreign Direct Investment." *The Review of Economics and Statistics* 78: 181 86. Uses data from a survey of 100 U.S. firms in 1991 to assess the importance of IPRs for investment in 14 countries.

Mansfield, Edwin. 1994. "Intellectual Property Protection, Foreign Direct Investment, and Technology Transfer." Discussion Paper No. 19. Washington, DC: Interna tional Finance Corporation. Survey evidence from 100 U.S. firms showing that IPRs influence FDI decisions only for certain industries.

Maskus, Keith E., Kamal Saggi, and Thitima Puttitanun. 2005. "Patent Rights and International Technology Transfer through Direct Investment and Licensing." In *International Public Goods and Transfer of Technology under a Globalized Intellectual Property Regime,* edited by Keith E. Maskus and Jerome H. Reichman. Cambridge: Cambridge University Press, 265 81. A theoretical model and empirical evidence showed the tradeoff be tween FDI and licensing.

Primo Braga, Carlos A., and Carsten Fink. 1998. "The Relationship between Intellectual Property Rights and Foreign Direct Investment." *Duke Journal of Compara tive and International Law* 9: 163 87. An often cited article reviewed implications of IPRs on FDI and pre sented empirical analysis that showed no significant re lationship between the two.

Rapp, Richard, and Richard Rozek. 1990. "Benefits and Costs of Intellectual Property Protection in Developing Countries." *Journal of World Trade* 24: 75 102. The first paper to develop an IPRs index rating the patent laws.

Smarzynska, Beata. 2004. "The Composition of Foreign Direct Investment and Protection of Intellectual Prop erty Rights: Evidence from Transition Economies." *European Economic Review* 48: 39 62. Examines whe ther IPRs affect composition of FDI flows for 24 tran sition countries in 1995.

Smith, Pamela J. 2001. "How Do Foreign Patent Rights Affect U.S. Exports, Affiliate Sales, and Licenses?" *Journal of International Economics* 55: 411 40. First empirical analysis that takes into account the simulta neous decision of multinational firms in choosing an entry mode.

Yang, Guifang, and Keith E. Maskus. 2001. "Intellectual Property Rights, Licensing, and Innovation in an En dogenous Product Cycle Model." *Journal of Interna tional Economics* 53: 169 87. The theoretical model showed importance of patent laws in reducing the li censing cost.

THITIMA PUTTITANUN

■ **Inter-American Development Bank**
See regional development banks

■ **interest parity conditions**
Interest parity conditions are no-arbitrage profit conditions for financial capital. When such condi tions hold, it is infeasible for investors to obtain higher returns by borrowing or lending. Hence, in principle, interest parity conditions define theoretical linkages between interest rates and exchange rates between countries.

The easiest way to understand parity conditions is to consider how a typical investor can save in dif ferent locations. Suppose the home currency is a dollar, and the foreign currency is a euro. Further assume that a forward market exists. A forward

contract allows an investor to enter into an agreement this period to exchange currencies k periods hence at a forward rate F known today. Then the investor can either save at home, receiving interest rate i, or save abroad, converting by the exchange rate S, receiving the foreign interest rate i^*, and then converting back to home currency by the forward rate F obtaining at time t for a trade at time $t+1$.

$$(1+i) \qquad versus \qquad (1+i_t^*) \times \frac{F_{t,t+1}}{S_t}$$

If the gross return on the left is greater than that on the right, then investors will place their capital in the home country; if it is less, then investors will place their capital abroad. With infinite amounts of capital moving in search of the highest return (and in this example, there is no risk in nominal terms), these returns will be equalized.

$$(1+i) = (1+i_t^*) \times \frac{F_{t,t+1}}{S_t} \qquad (1)$$

After manipulation,

$$\frac{(i-i^*)}{(1+i_t^*)} = \frac{F_{t,t+1} - S_t}{S_t} \qquad (2)$$

This condition is called "covered interest rate parity," reflecting the fact that investors are "covered" against nominal uncertainty by way of the forward market.

If the forward rate is equal to the future spot rate, such that $F_{t,t+1} = S_{t,t+1}^e$, then (2) becomes:

$$\frac{(i-i^*)}{(1+i_t^*)} = \frac{S_{t,t+1}^e - S_t}{S_t} \qquad (3)$$

where the e superscript denotes "expected." Equation (3) is termed *uncovered interest rate parity*. This expression holds when investors do not require compensation for the uncertainty associated with trading currencies in the future. It states that *expected* nominal returns are equalized across borders in common currency terms.

When interest rates are low, the following log approximations are often used for equations (2) and (3).

$$(i_t - i_t^*) = f_{t,t+1} - s_t \qquad (2')$$
$$(i_t - i_t^*) = s_{t,t+1}^e - s_t \qquad (3')$$

where f and s are the logs of the forward and spot rates, respectively.

Frankel (1991) has labeled condition (2) holding as characterizing perfect capital mobility, while condition (3) is associated with perfect capital substitutability. These terms arise from the view that if (2) does not hold, there must be some sort of impediment capital controls or the threat thereof to the free flow of financial capital. But even if capital is free to move, investors may still respond to risk; that response to risk might drive a wedge between the expected spot and forward rates. When investors are risk-neutral in nominal terms, then investors will treat capital (say debt instruments issued in different currencies) as perfectly substitutable.

The conditions just discussed pertain to financial capital. In order to consider the mobility of physical capital, one has to bring into play the prices of commodities. Integration of goods markets are often defined as relative purchasing power parity (PPP) holding. Ex ante relative PPP can be written as:

$$s_{t,t+1}^e - s_t = (p_{t,t+1}^e - p_t) - (p_{t,t+1}^{*e} - p_t^*) \quad (4)$$

where p is the log price level. Equation 4 states that expected depreciation equals the expected inflation differential. Combining (4) with the uncovered interest rate parity condition (3′) leads to real interest parity.

$$i_t - (p_{t,t+1}^e - p_t) = i_t^* - (p_{t,t+1}^{*e} - p_t^*) \quad (5)$$

This says that the expected rate of return on capital, expressed in physical units, is equalized across borders. To the extent that in neoclassical models the marginal product of capital equals the real interest rate, this condition is equivalent to the equalization of marginal product of capital equalized across borders.

Covered Interest Parity Assessed For developed economies since the dismantling of capital controls, covered interest parity holds fairly well. It should be noted that most tests are conducted using offshore rates, in which case (2) is sometimes termed *closed interest parity*, although the term *covered interest parity* is often used to encompass this concept.

Early tests were conducted by Frenkel and Levich (1975). They found that, after accounting for transactions costs, covered interest parity held for

three-month horizons. Offshore rates sometimes diverge from onshore rates, so that the findings of covered interest parity are somewhat weaker.

The question of whether covered interest parity holds for longer horizons is an open one. Popper (1993) concludes that covered interest differentials at long maturities are not appreciably greater than those for short (up to one year) maturities. This is a surprising result given that there are likely a number of regulatory impediments that would tend to introduce frictions into the arbitrage process.

Prior to the dismantling of capital controls, and in many emerging markets today, covered interest parity is unlikely to hold. In other words, covered interest differentials could be interpreted as political risk, associated with the possibility of governmental authorities placing restrictions on deposits located in different jurisdictions (clearly this is not relevant when all the deposits are offshore). Aliber (1973) is credited with this interpretation, while Dooley and Isard (1980) provided empirical estimates for the deutsche mark/dollar rate.

The Empirical Evidence for Uncovered Interest Parity Uncovered interest parity (UIP) is a more difficult condition to test, essentially because expected exchange rate changes are unobservable. In the literature, most tests of UIP are actually joint tests of UIP and the rational expectations hypothesis, that is, that ex post realizations of the exchange rate are an unbiased measure of the ex ante exchange rate, namely, $s_{t,t+1}^e = E(s_{t+1}|I_t)$. This assumption combined with equation (2′) yields this standard regression equation, sometimes called the Fama equation (Fama 1984):

$$s_{t+1} - s = \beta_0 + \beta_1(f_{t,t+1} - s_t) + v_{t+1} \qquad (6)$$

Or by virtue of covered interest parity holding,

$$s_{t+1} - s = \beta_0 + \beta_1(i_t - i_t^*) + v_{t+1} \qquad (7)$$

where under the joint null hypothesis v_{t+1} is a mean zero error unpredictable using past information, and $\beta_1 = 1$.

The evidence in favor of this joint hypothesis of UIP and rational expectations is quite weak. The regression of the ex post change of the spot exchange rate on either the forward discount (in equation 6) or the interest differential (in equation 7) typically

yields a slope coefficient estimate that is not only different from unity, but in fact negative and different from zero at conventional levels of statistical significance. This is true for reserve currencies (the U.S. dollar, the yen, the Swiss franc, the deutsche-mark, the franc, or the euro) at horizons up to a year. It is also true for some emerging market currencies (see Frankel and Poonawala 2006). One interesting characteristic of these regressions is that, although the coefficients are typically different from zero in a statistical sense, the proportion of total variation explained is typically very small.

At longer horizons (3, 5, and 10 years) the evidence is more supportive of the combined UIP rational expectations hypothesis. Chinn and Meredith (2004) document that estimates of the β_1 coefficient are usually not significantly different from the posited value of unity at 5- and 10-year horizons. The finding that the joint hypothesis of UIP and rational expectations holds better at long horizons than at short appears to be robust. Nonetheless, some caution is necessary here. Consider regressions involving 10-year interest differentials; by 2003, there would be only three nonoverlapping observations available per currency. Interestingly, Chaboud and Wright (2005) find that UIP also holds at extremely short horizons of a few minutes.

Other interesting results pertain to periods of extreme market turmoil. Flood and Rose (2002), following their 1996 work, find that UIP holds better in recent times when the sample encompasses successful attacks on currency pegs. Nonetheless, they still find lots of heterogeneity in experiences with UIP.

A different perspective on UIP is provided by dropping the rational expectations hypothesis.

A new area of research involves an investigation of whether UIP holds for emerging markets. Bansal and Dahlquist (2000) found that there was a basic asymmetry in whether UIP holds. In particular, they find that when the U.S. interest rate is lower than foreign country rates, UIP holds, while UIP fails to hold when the U.S. rate is higher. They also find that idiosyncratic factors, such as the gross domestic product per capita of the foreign country, are

important in determining the degree of failure of UIP to hold.

Using the forward discount instead of interest differentials, Frankel and Poonawala (2006) find that there is substantial heterogeneity in the results. What matters importantly is the exchange rate regime: highly managed exchange rate regimes are associated with currencies that exhibit greater deviations from UIP.

Real Interest Parity Measured If UIP does not seem to hold at short horizons, it seems unlikely that real interest parity, described as exact equalization of real interest rates, would hold. One could, however, still test the weaker condition that movements in real rates in one country would be met by one-for-one real movements in other countries.

The key difficulty with testing this condition, like that with testing UIP, is that market expectations are not directly observable. Hence, one can conduct only joint tests for real interest parity. In Fujii and Chinn (2001), real interest rates are calculated using a variety of proxy measures of expected inflation: ex post inflation, and inflation predicted using lagged values of inflation models. Both approaches are consistent with rational expectations. They find that the real interest parity holds with different strength at different horizons. As in numerous previous studies (Cumby and Obstfeld 1984; Mark 1985), the real interest parity (RIP) hypothesis is decisively rejected with short horizon data. At 5- to 10-year horizons, however, the empirical evidence becomes far more supportive and in some cases the RIP hypothesis is not rejected. In general, RIP, *up to a constant*, holds better at long horizons than at short horizons. These results are robust to alternative ways of modeling expected inflation rates.

In recent years, several countries, including the United Kingdom, the United States, France, and Canada, have begun issuing inflation-indexed debt securities. These are marketable securities whose principal is adjusted by changes in the price level (usually the consumer price index). The principal increases with the inflation rate so that the real return can be directly observed. A cursory investigation re-

veals that there is no evidence of equalization. Moreover, although there is some covariation, it is not anywhere near one for one. The thinness of the markets and the differences in the maturities of the relevant debt instruments, however, make strong conclusions in either direction difficult.

See also capital mobility; exchange rate regimes; exchange rate volatility; forward premium puzzle; peso problem; purchasing power parity; reserve currency; sovereign risk; speculation

FURTHER READING

Aliber, Robert. 1973. "The Interest Parity Theorem: A Reinterpretation." *Journal of Political Economy* 81: 1451 59. An early formalization of the idea that political risk the possibility of the imposition of capital controls or regulations might drive a wedge between observed returns.

Bansal, Ravi, and Magnus Dahlquist. 2000. "The Forward Premium Puzzle: Different Tales from Developed and Emerging Economies." *Journal of International Economics* 51: 115 44. Documents some stylized facts regarding interest rate differentials and changes in exchange rates for a wide set of currencies.

Chaboud, Alain, and Jonathan Wright. 2005. "Uncovered Interest Parity: It Works, but Not for Long." *Journal of International Economics* 66 (2): 349 62. Demonstrates that uncovered interest rate parity holds at extremely short horizons of several minutes.

Chinn, Menzie D. 2006. "The (Partial) Rehabilitation of Interest Rate Parity: Longer Horizons, Alternative Expectations and Emerging Markets." *Journal of International Money and Finance* 25 (1) (February): 7 21. Surveys the empirical literature regarding uncovered interest rate parity.

Chinn, Menzie D., and Guy Meredith. 2004. "Monetary Policy and Long Horizon Uncovered Interest Parity." *IMF Staff Papers* 51 (3) (November): 409 30. Documents the tendency for interest rate differentials to be better predictors of subsequent exchange rate movements at longer horizons, and to provide a theoretical model based on central bank behavior.

Cumby, Robert E., and Maurice Obstfeld. 1984. "International Interest Rate and Price Level Linkages under Flexible Exchange Rates: A Review of Recent Evidence."

In *Exchange Rate Theory and Practice*, edited by J. F. O. Bilson and R. C. Marston. Chicago: University of Chicago Press, 121 51. Provides an early test of real interest parity, using the rational expectations methodology.

Dooley, Michael, and Peter Isard. 1980. "Capital Controls, Political Risk, and Deviations from Interest rate Parity." *Journal of Political Economy* 88 (2): 370 84. Gives empirical estimates in support of the proposition that deviations from interest rate parity may be caused by the presence of capital controls.

Fama, Eugene F. 1984. "Forward and Spot Exchange Rates." *Journal of Monetary Economics* 14 (3): 319 38. Provides early estimates of canonical regressions of exchange rate changes on the forward discount.

Flood, Robert, and Andrew Rose. 1996. "Fixes: Of the Forward Discount Puzzle." *Review of Economics and Statistics* 78 (4): 748 52. Documents the fact that when there are expectations of large exchange rate changes as during currency crises interest differentials do predict the direction of subsequent exchange rate changes.

. 2002. "Uncovered Interest Parity in Crisis." *IMF Staff Papers* 49: 252 66. Documents the fact that when there are expectations of large exchange rate changes as during currency crises interest differentials do predict the direction of subsequent exchange rate changes.

Frankel, Jeffrey A. 1991. "Quantifying International Capital Mobility in the 1980s." NBER Working Paper No. W2856. Available at http://ssrn.com/abstract= 284049. Systematically measures the extent of capital mobility and makes a distinction between perfect capital mobility and perfect capital substitutability.

Frankel, Jeffrey, and Jumana Poonawala. 2006. "The Forward Market in Emerging Currencies: Less Biased Than in Major Currencies." NBER Working Paper No. 12496 (August). Cambridge, MA: National Bureau of Economic Research. Shows the predictive characteristics of the forward discount for emerging market currencies.

Frenkel, Jacob, and Richard Levich. 1975. "Covered Interest Arbitrage: Unexploited Profits?" *Journal of Political Economy* 83 (2): 325 38. Examines a transactions cost based explanation for why covered interest parity does not hold exactly.

Froot, Kenneth, and Jeffrey Frankel. 1989. "Forward Discount Bias: Is It an Exchange Risk Premium?" *Quarterly Journal of Economics* 104 (1) (February): 139 61. Uses survey data on exchange rate expectations to determine whether a risk premium or biased expectations explain the forward premium puzzle.

Fujii, Eiji, and Menzie D. Chinn. 2001. "Fin de Siècle Real Interest Parity." *Journal of International Financial Markets, Institutions and Money* 11 (3/4): 289 308. Uses several methods of measuring expected inflation to test whether real interest rates are equalized across borders.

Mark, N. C. 1985. "Some Evidence on the International Inequality of Real Interest Rates." *Journal of International Money and Finance* 4: 189 208. Provides an early test of real interest parity.

Popper, Helen. 1993. "Long Term Covered Interest Parity Evidence from Currency Swaps." *Journal of International Money and Finance* 12 (4): 439 48. Examines whether covered interest parity holds at multi year horizons.

MENZIE D. CHINN

■ internalization theory

Internalization refers to the decision by a multinational firm producing abroad whether to own a foreign production facility or to license or contract with a foreign firm to produce a good or service on behalf of the multinational. Specifically, internalization refers to production internal to the ownership boundaries of the firm. This topic has a long history in the international business literature, but has more recently emerged under the converse name of outsourcing. This is merely a change in name: the firm must decide whether to internalize or outsource. Other terms used to refer to this decision include *mode choice* and *boundaries of the firm.*

This entry focuses on relatively recent papers that offer alternative formal models of the internalization decision. Nevertheless, it is important for researchers in this field to acknowledge the substantial contributions of earlier writers in the international business tradition, including John Dunning, Mark Casson, Alan Rugman, Oliver Williamson, and others. An excellent review of this literature is provided in Caves (2007) and in a somewhat more circumscribed fashion in Markusen (2002).

Firms make mode choices, such as among exports, subsidiary production, and outsourcing, in response

to the conditions and constraints they experience. A typical starting point in the theory of the multinational is that the firm possesses proprietary assets that give it a motive for expanding abroad in the first place. These assets could include product or process technology, a brand name or trademark, or a reputation for product quality. But in moving or using these assets abroad, the firm runs certain risks of asset dissipation or holdup through opportunistic behavior of foreign licensees, contractors, or even by the firm's own managers in the case of an owned subsidiary. Asset dissipation is therefore a motive for internalization. But this must be balanced against other competing forces, such as giving the local manager or licensee strong incentives to engage in efficient investment and effort levels. Internalization theory explores the tension among these alternative factors and derives the optimal mode choice for firms subject to various assumptions about contracting and other constraints.

The Internalization Decision and Opportunistic Behavior Some of the first formal models of the internalization decision were published in the late 1980s and 1990s and draw their empirical motivation from the strong association of multinationality with knowledge-based assets such as those described in the previous paragraph (again, see Caves 2007 and Markusen 2002). On the one hand, these assets (or the services thereof) are easily transferred overseas, such as providing a blueprint, chemical formula, or procedure to a foreign plant. On the other hand, the same characteristics that make it easy to transfer these assets make them easily learned by foreign managers, agents, or licensees. Once the agent sees the blueprint or license he or she could defect to produce the product in a new firm.

Horstmann and Markusen (1987) have a model in which the intangible or knowledge-based asset is a reputation for product quality. This quality can only be observed by a consumer after purchase and use of the product. The difficulty is that, if the firm extracts all of the rents from a local licensee, that licensee has an incentive to product a cheap, low-quality substitute and earn positive profits for one period before being fired. The multinational risks asset dissipation

through agent opportunism, a theme that occurs repeatedly in subsequent papers.

The Horstmann-Markusen (1987) paper tends to focus on a firm's decision to export versus produce abroad. It is less well suited to explicitly considering the decision to create an owned subsidiary versus using a foreign licensee; after all, the local manager has much the same incentives as a licensee. Two 1996 papers made some further progress on this issue. Ethier and Markusen (1996) have a model in which a firm introduces a new product every second period. During the first period of a product's life, the local agent or manager absorbs all relevant aspects of the technology and is capable of producing the product on his or her own in the second period. At the end of the second period, the product is obsolete and the manager's knowledge is of no further use.

Under certain parameterizations of this model, the multinational can extract all rents with a single-product licensing contract, and so this is of course optimal. For other parameterizations, this cannot be done. The alternative is for the multinational to offer a contract that shares some rents with the local manager, and this can be credible and incentive-compatible for both parties over the long term in which the contract and rent sharing is repeated with every new product. This long-term incentive-compatible sharing arrangement is thought of as a subsidiary relationship, not limited to one product but continuing indefinitely into the future.

Horstmann and Markusen (1996) have quite a different approach, introducing asymmetric information along with moral hazard. The potential local agent or manager possesses local information on market demand or costs that is not known to the multinational. The market could actually be large, but the agent has an incentive to pretend it is small and shirk on effort so that sales are consistent with high effort in a small market. The best feasible licensing contract forces the multinational to share rents, termed information rents, with the agent. The alternative is for the firm to establish an owned subsidiary from the beginning, which is more costly, but the firm learns the true state of demand or costs without having to share rents to extract this infor-

mation. In the vernacular, an owned subsidiary gives the multinational a larger share of a smaller pie, another theme that occurs repeatedly in the internalization/outsourcing literature.

Three later papers followed in the tradition of the Ethier and Markusen (1996) model: Fosfuri, Motta, and Rønde (2001); Glass and Saggi (2002); and Markusen (2001). All share the idea that knowledge is absorbed by workers in the course of production, and they are then at least potentially able to defect and start rival firms after some period of time. The Fosfuri, Motta, and Rønde model has a richer technological structure than the original Ethier and Markusen model. In particular, the managers or licensees who absorb the knowledge from the multinational may later enter into a complementary relationship with their former employer rather than a competitive one, supplying goods or services to the multinational. This is a valuable development insofar as there are many examples of former employees, in Taiwan, for example, becoming contractors in a cooperative relationship with their old firms.

Glass and Saggi (2002) focus on workers in general and not managers or licensees specifically. Workers absorb knowledge and can be hired away by local firms. The multinational has two options to prevent the dissipation of its technology, termed a transfer of technology (i.e., the term *technology transfer* in this model refers to transfer from the multinational to the host-country firm, not to whether or not the multinational uses or "transfers" its technology to the host-country plant). The firm can pay its workers a wage premium, or it can choose an alternative, more costly location (e.g., export to the host country). Glass and Saggi solve for parameters under which the multinational prevents technology transfer by paying sufficiently high wages and under which the multinational's best option is to pay minimal wages and accommodate the technology transfer through worker mobility. The welfare consequences of these alternatives are analyzed. There is no explicit modeling of a licensing or outsourcing alternative to establishing a subsidiary.

Markusen (2001) follows in the Ethier-Markusen tradition as just noted. He extends this model to in-

clude an explicit analysis of intellectual properties rights and provides a welfare analysis. Tighter intellectual property protection and contract enforcement raise the costs to the local manager of defecting, but also raise the costs to the multinational of firing and replacing the manager after one period. Tighter enforcement means that the incentive-compatible contract allows the multinational to share fewer rents with the local manager or licensee. This presents a trade-off for the host country in setting policy. Tighter enforcement means that the multinational may now enter the country whereas before the optimal policy was to export, and this is welfare improving for the host. On the other hand, for a firm that is entering anyway, tighter enforcement transfers rents from the local manager to the firm and hence is welfare reducing.

Property Rights Approach and the Holdup Problem About the same time as this last set of papers, an important advancement to the internalization question was being developed by Antrás, Grossman, and Helpman (Antrás 2003, 2005; Antrás and Helpman 2004; Grossman and Helpman 2004). All of these authors substitute the term *outsourcing* for the term *internalization*. Their approach is sometimes termed the "property-rights" approach to the firm after contributions by Grossman and Hart (1986) and Hart and Moore (1990), following earlier contributions of Williamson (1985) and others in the international business literature.

The new literature combines a number of separate elements that together produce a coherent model that offers clear empirical predictions. The first element (in no particular order) is the assumption that production requires "relation-specific investments," meaning that a multinational and a foreign individual or firm must incur sunk investments prior to production that have no outside value if the relationship breaks down. The second element is the assumption of incomplete contracting: certain things are simply not contractible or alternatively any contract on these items is not enforceable. The papers just mentioned focus on the idea of noncontractibility, in particular they assume that certain things either cannot be observed and/or are not verifiable by third parties rather than focusing on institutional failures such that contracts cannot be en-

forced. These two elements of the newer theory are shared with the older literature reviewed above.

The assumptions of sunk investments and non-contractibility lead to a third problem, which is ex-post "holdup." What happens after production occurs cannot be contracted ex ante, and so each party has some ability to negotiate ex post and to prevent the other party from fully utilizing the output. As is generally assumed in this approach to the firm, the Antrás-Grossman-Helpman papers assume that the multinational and the local manager or firm engage in ex post Nash bargaining over the surplus. The final element of this approach is the notion of ownership, necessary in order to distinguish an owned subsidiary from an arm's-length supplier. The property rights approach defines the owner of the foreign production facility as the party that has residual rights of control; that is, the party that owns any output or other assets of the firm in the event that bargaining breaks down.

The difficulty for the multinational is that the choice between an owned subsidiary and an arm's-length licensee both entail some inefficiency again due to incomplete contracting. Antrás (2003), in an elegant model and exposition, assumes that the multinational chooses the quantity of capital to be used in production and the local manager or licensee chooses the quantity of labor. In the Nash bargaining phase, the problem is that each party will receive only a fraction of the marginal return to its ex ante investment, and thus there will be investment in both capital and labor (in Antrás's model) in equilibrium. Ownership of the local supplier entitles the multinational to residual rights of control and thus improves the ex post bargaining position, but reduces the local manager's incentive to make an efficient ex ante investment.

Antrás assumes that there are many potential host-country licensees or managers and that the multinational can impose an up-front fee and so competition among suppliers will lead the up-front fee to adjust so that the local supplier will just break even. Let X be the output that is produced. Because X is assumed to be a component that is specific to the multinational firm, it also follows that the local manager/licensee has no ability to use X outside the relationship even if he or she

has residual ownership rights, and so the manager/licensee's outside option is zero under either the subsidiary or licensing relationship. But the multinational's outside option is higher with residual rights of control. The firm can use at least part of the X when the multinational is the owner, but neither party can use X in the licensing relationship when bargaining breaks down. Thus the multinational will capture a larger share of the surplus in ex post bargaining under a subsidiary arrangement.

The consequence when combining these various elements is that, while both a subsidiary and a licensing relationship suffer from underinvestment, the subsidiary relationship suffers more from an underinvestment in labor relative to capital: the licensee has less incentive to invest due to a smaller marginal share. The licensing arrangement suffers more from an underinvestment in capital, since the multinational receives a smaller share in the ex post bargaining outcome. The next step becomes straightforward: the subsidiary is the more profitable option for a capital-intensive industry while the arm's-length contract is more profitable for a labor-intensive industry. This clear prediction is successfully taken to the data.

A conceptually related model is found in Antrás (2005). A product-cycle model is developed, in which a high-tech input that can be produced only in the North is combined with a low-tech input that can be produced either in the North or South. As time passes, the technology shifts so that the high-tech input is less important. The theory predicts that, as time passes, production first shifts to the South within firm boundaries and only later to independent foreign firms.

Various extensions and refinements of this approach are found in other Antrás-Grossman-Helpman papers noted above. These are an advance on the earlier literature discussed above, insofar as the difference between an owned subsidiary and an arm's-length relationship is sharper and better defined. Empirical predictions are similarly sharper and then have so far stood up under empirical investigation such as in Feenstra and Hanson (2005) in addition to Antrás's own results.

Chen, Horstmann, and Markusen (2007) have produced a new model in which there is both knowl-

edge capital and physical capital. As in the earlier literature, the knowledge capital is absorbed by the workers over the first period and hence there is no clear ownership issue: knowledge is nonexcludable. The physical capital is fully excludable as in the newer property-rights papers and belongs to the residual claimant, the owner of the production facility. They obtain the prediction that the owned subsidiary is more likely to be chosen in an industry in which the ratio of knowledge capital to physical capital is high. They suggest that this could be proxied by a Tobin's Q statistic in subsequent empirical analysis. The results seem consistent with the extensive empirical analysis surveyed in Caves (2007).

Areas Needing More Research All of these models are helpful in understanding the internalization versus outsourcing decision, and they seem to be standing up well to empirical analysis as noted. They are, however, very specific models. The international business literature, while less formal and analytic, identifies a far broader range of factors influencing a firm's entry mode choice in serving a foreign market. Judging by the references contained in most of the papers surveyed here, the international business literature under the term *internalization* is generally being ignored in the literature under the name *outsourcing*. This is surely a mistake, and researchers in the more formal economics tradition are missing out on the richness of detail, ideas, and empirical evidence found in the international business literature. Again, Caves (2007) is an excellent background for any scholar interested in this field.

We can conclude then with a brief listing of some of the other factors that seem important and which are ripe pickings for more formal analysis. Complex tax systems influence firms' mode choices. Formal barriers to majority-owned affiliates, often indirect and hard to measure, corrupt observations from developing countries and suggest more joint ventures and arm's-length arrangements than would actually be observed in less-regulated markets. A broad range of institutions including legal systems, contract enforcement, corrupt practices, host government holdup, and intellectual property protection influence the relative attractiveness of owned affiliates versus licensing and other hybrid mode forms.

Some disconnect with empirical evidence is also evident. While many of the newer formal models surveyed here assume the foreign partner is providing an intermediate for the multinational, empirical evidence consistently shows that the major motive for foreign direct investments is to serve local markets, with production for regional third countries coming in second. Production for export back to the parent country amounts to only about 10 percent of sales for foreign affiliates of U.S. manufacturing multinationals and even then, much of this small amount is final goods assembled in labor-intensive factories abroad. Examinations of lengthy, complex outsourcing contracts apparently cast some doubt on convenient assumptions that a whole range of things are noncontractible since they appear very explicitly in these real-world contracts, but there is little published evidence on this point.

Other even more intangible but no doubt very real considerations are often mentioned in the international business literature. Dissimilar corporate cultures and conflicts between the global objectives of the multinational and the more circumscribed and short-run objectives of the potential local contracting firm are just two of many examples. Multinationals are not trying to maximize local profits, and the task of inducing a local affiliate, much less a licensee, to respond to the firm's global strategy is a difficult contracting complication. All of these factors suggest that there is still much to be done.

See also foreign direct investment: the OLI framework

FURTHER READING

Antrás, Pol. 2003. "Firms, Contracts, and Trade Structure." *Quarterly Journal of Economics* 118 (4): 1375 1418. Shows that the subsidiary option is more profitable for capital intensive industries, whereas arm's length contracting is more profitable for labor intensive industries.

———. 2005. "Incomplete Contracts and the Product Cycle." *American Economic Review* 95 (3): 1054 71. Product cycle model in which manufacturing is shifted abroad first within firm boundaries, and at a later stage to independent foreign firms.

Antrás, Pol, and Elhanan Helpman. 2004. "Global Sourcing." *Journal of Political Economy* 112 (3): 552 80. Extension/refinement of Antrás (2003) to include varying degrees of contractibility across inputs and countries in a model of whether to integrate into the production of intermediate inputs or outsource them, and from which country to source the inputs.

Caves, Richard E. 2007. *Multinational Enterprise and Economic Analysis*. 3d ed. Cambridge: Cambridge University Press. Excellent background reading on international business approach.

Chen, Yongmin, Ignatius J. Horstmann, and James R. Markusen. 2007. "Physical versus Knowledge Capital as Determinants of Multinational Firms' Mode Choices for Foreign Production." University of Colorado Working Paper. Boulder, CO. Predicts that subsidiaries should be more likely in industries with a high ratio of knowledge capital to physical capital.

Ethier, Wilfred J., and James R. Markusen. 1996. "Multinational Firms, Technology Diffusion and Trade." *Journal of International Economics* 41 (1 2): 1 28. Model of contracting issues, division of rents, and mode choice when manager/licensee learns all about technology and becomes potential rival.

Feenstra, Robert C., and Gordon H. Hanson. 2005. "Ownership and Control in Outsourcing to China: Estimating the Property Rights Theory of the Firm." *Quarterly Journal of Economics* 120 (2): 729 61. Studies factory ownership and input control for trade processing in China, where foreign firms contract out assembly of intermediates into final products.

Fosfuri, Andrea, Massimo Motta, and Thomas Rønde. 2001. "Foreign Direct Investments and Spillovers through Workers' Mobility." *Journal of International Economics* 53 (1): 205 22. Studies conditions under which technology spillovers occur to workers trained with a superior technology leaving to work for a local competitor.

Glass, Amy J., and Kamal Saggi. 2002. "Multinational Firms and Technology Transfer." *Scandinavian Journal of Economics* 104 (4): 495 513. Model of host country policy toward FDI when firms may decide to pay a wage premium or to produce elsewhere, in order to protect their technology from leaving to rival rivals through the movement of workers.

Grossman, Gene M., and Elhanan Helpman. 2004. "Managerial Incentives and the International Organization of Production." *Journal of International Economics* 63 (2): 237 62. Extension/refinement of Antrás (2003): models location and organization choices of heterogeneous firms and how these are affected by falling trade costs.

Grossman, Sanford J., and Oliver D. Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy* 94 (4): 691 719. A first paper developing the property rights approach to the firm.

Hart, Oliver D., and John Moore. 1990. "Property Rights and the Nature of the Firm." *Journal of Political Economy* 98 (6): 1119 58. A second paper, drawing on Grossman and Hart (1986), on the property rights approach to the firm.

Horstmann, Ignatius J., and James R. Markusen. 1987. "Licensing Versus Direct Investment: A Model of Internalization by the Multinational Enterprise." *Canadian Journal of Economics* 20 (3): 464 81. Models contracting issues related to product quality rents must be shared to provide licensee incentive not to defect by selling low quality.

———. 1996. "Exploring New Markets: Direct Investment, Contractual Relations, and the Multinational Enterprise." *International Economic Review* 37 (1): 1 20. Model of mode choice emphasizing the trade off between information rents under licensing versus higher costs of establishing a subsidiary.

Markusen, James R. 2001. "Contracts, Intellectual Property Rights, and Multinational Investment in Developing Countries." *Journal of International Economics* 53 (1): 189 204. Captures how intellectual property rights and contract enforcement affect the division of profits between a multinational and local manager/licensee.

———. 2002. *Multinational Firms and the Theory of International Trade*. Cambridge, MA: MIT Press. Formal theory and empirical estimation of MNE models; concentrates somewhat on defining, developing, and estimating the knowledge capital model.

Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism*. New York: Free Press. A classic work on the boundaries of the firm.

JAMES R. MARKUSEN

■ international financial architecture

The international financial architecture is the set of institutions and norms both international and domestic that shape the international financial system. The institutional relations among industrial economies are clearly part of the international financial architecture. The policy debate that followed the currency crises in Asia and other emerging market crises in the late 1990s focused on reforms that could help make the integration of emerging economies into the international financial system less disruptive.

The resulting debate rested on two assumptions. First, private capital should normally flow from advanced economies to emerging economies at least to those with sound economic fundamentals and solid policy frameworks. Second, the integration of emerging economies into global financial markets a process that was catalyzed by the 1989 90 Brady plan, which transformed bad loans left over from the 1980s debt crisis into bonds had not proceeded as smoothly as many had hoped. The "architects" consequently sought to create an institutional structure that would sustain private financial flows to emerging economies while reducing the risk that sudden interruptions in these flows would lead to painful crises and also increasing the international community's ability to manage crises that could not be avoided.

Background U.S. treasury secretary Robert Rubin first used the term *international financial architecture* in a speech in April 1998. The resulting debate culminated in the Cologne communiqué in 1999. It emphasized the need for a higher standard of transparency, stronger financial systems in emerging economies, changes in financial regulation in industrial economies to help make capital flows less cyclical, "corner" exchange rate regimes, and greater efforts to involve private creditors in the provision of crisis financing. This agenda informed much of the work of the International Monetary Fund (IMF) over the next few years.

It would be a mistake, however, to limit analysis of the architecture reforms solely to the specific proposals that followed from Rubin's speech: the debate over the need to reform the international financial architecture started well before Rubin coined the term *architecture* and did not end when the term disappeared from the official sector's formal lexicon in 2001. After Mexico's crisis of 1994 95, the Group of Seven (G7) called for the development of new facilities designed to allow the IMF to lend larger sums, but for shorter periods and at higher rates, as well an expansion of the IMF's lending capacity. These reforms were implemented in 1997 and 1998 with the creation of the Supplemental Reserve Facility and congressional approval of an IMF quota increase. The substantive debate over many key issues continued through 2003, when Mexico's decision to introduce collective action clauses (CACs) provisions that allowed a supermajority of bondholders to amend a bond's financial terms ended the debate over IMF's proposal for a new international sovereign bankruptcy regime. The intensity of the debate over the right scale of IMF lending waned in the absence of any major crises. New topics, notably a surge in the U.S. current account deficit largely financed by a rise in the current account surplus of the emerging world, moved to the top of the international economic and financial policy agenda.

During the peak of the debate over financial crises in emerging economies, many complained that talk of architecture was too grand and that "plumbing" would be a more appropriate metaphor. Such arguments came from two directions: those who thought better plumbing was all that the international financial system needed, and those who thought the G7 was making a mistake by just fiddling with the plumbing when a truly new architecture was needed.

Ambitious Proposals The Asian crisis certainly generated a host of ambitious proposals. The IMF's first deputy managing director Anne Krueger proposed radically changing the institutions for debt restructuring by amending the IMF's Articles of Agreement to provide sovereign governments in default with bankruptcy-style protections (the sovereign debt restructuring mechanism, or SDRM). The economist Joseph Stiglitz suggested a "super" Chapter 11 regime to facilitate an across-the-board restructuring of private borrowers' debts in the event of macroeconomic shocks.

Anne Krueger's predecessor, Stanley Fischer, focused more on the institutional structure for crisis lending than on the institutional structure for debt restructuring. Fischer argued that the IMF should be transformed into a lender of last resort, able to lend in quantities that would be sure to end cross-border runs. The International Financial Institutions Advisory Committee (2000), more commonly called the Meltzer Commission, suggested that the IMF get out of the business of lending to countries that promise to deliver sound macroeconomic policies only when they are close to default and instead lend large sums to countries that qualified in advance for extra protection by maintaining good policies. The International Financial Institutions Advisory Committee (IFIAC) specifically recommended that the IMF lend large sums for short-term loans   120 days with only one possible rollover   to countries that prequalified for support. The IFIAC also suggested that the strength of a country's banking system be the key criterion for determining eligibility. Others have proposed less draconian forms of prequalification.

Proposed changes in governance of the major international financial institutions were equally dramatic, with proposals tabled to merge the IMF and the World Bank, to eliminate both institutions, and to create a new global financial regulator.

More Modest Results In the end, calls for major reforms generally were rejected in favor of more incremental changes. The IMF toyed with a facility based on prequalification (the contingent credit line, or CCL) but never gave up its traditional crisis lending. A new forum that brought regulators together with finance ministries, central banks, and the IMF (the Financial Stability Forum) and the development of new international codes and standards substituted for the creation of an international superregulator. No major reforms were made to the IMF's voting structure after the Asian crisis   though in 2006, the IMF did increase the quota of four underrepresented emerging economies: China, Korea, Mexico, and Turkey. The membership of informal clubs such as the G7 that often influenced IMF policy did not change, but a new forum, the G20, was created to bring the G7 together with major emerging markets.

Modest Reforms In the end, though, the architects did more than just clean up the plumbing. Some of the most consequential changes were adopted early on. After the crisis in Mexico, the United States decided to strengthen multilateral institutions for crisis financing. This led, in time, to important changes in the architecture. A bigger IMF was combined with facilities designed to allow the IMF to put more money on the table faster. In principle, exceptional IMF financing came with an expectation that those funds could be repaid more quickly than a typical IMF loan. In practice, the IMF's major shareholders did lend large sums to some countries that did not have a realistic chance of repaying the fund quickly; the large short-term loans provided to Turkey, Brazil, and Argentina in 2001 and 2002 all had to be refinanced. Large-scale crisis financing from the IMF effectively became the norm   despite claims by the G7 to the contrary   when major emerging economies encountered financial difficulties.

The institutions for debt restructuring evolved organically. The large international banks generally stopped providing medium- to long-term syndicated loans to the governments of emerging economies after the debt crisis of the 1980s, instead lending for shorter terms to private banks (some of which then lent to their local government) or private firms. Nonetheless, the architecture for coordinating the major international banks that was inherited from the sovereign debt crisis of the 1980s was adapted in Korea to help coordinate a rescheduling of interbank credit lines. New institutions for bond restructuring emerged, though not without some birth pangs. Negotiations between a distressed sovereign borrower and a coordinating committee of banks gave way to exchange offers. Distressed debtors, often advised by a major international bank, offered to exchange old bonds for new bonds with different payment terms. The success of these exchanges   participation was high and litigation proved less troublesome than initially anticipated   eliminated much of the pressure for more dramatic changes to the institutions for debt restructuring.

Arguably of most importance, the policies of most of the world's emerging economies changed dramatically. Emerging market economies embraced central bank independence and inflation targeting. Indeed, their inflation rates began to converge with those of the advanced countries. Local currency bond markets took off. Emerging economies concluded that they needed to hold far more reserves than they had held in the past; most now hold reserves well in excess of their short-term external debts. Many emerging economies began to allow their exchange rate to float, though most still intervene far more heavily than advanced economies. Transparency increased dramatically   at least for those emerging economies that felt compelled to live up to the IMF's revised standards.

Compared to the changes made in emerging markets, the changes in regulatory structure in industrial economies were quite modest. Concerns that capital flows from the industrial world to emerging economies were procyclical, with too much money flowing into emerging economies in good times and too much flowing out in bad times, did not prompt major policy changes. A revised set of Basel capital standards was in the process of being implemented as of 2007. Although hedge funds and the regulators of the banks that provide hedge funds with leverage were temporarily chastened by the 1998 implosion of a major U.S. hedge fund, Long-term Capital Management, the formal regulatory structure changed little. After a brief down period, hedge funds entered a new phase of phenomenal growth.

The debate on architecture ended with a broad consensus on the steps emerging economies needed to take to reduce the probability of crises, but with few tools that could force an emerging economy to follow these recommendations. A comparable consensus on the core issue of crisis resolution never truly emerged. The G7 was comfortable with an IMF that had the capacity to make big loans, and policy statements that claimed   without much credibility   that this capacity would not be used in the future. The G7   and for that matter the world's major emerging economies   never could define

when a debt restructuring should be an integral part of efforts to resolve a crisis.

Changes in the global economy, though, gradually shifted the focus of policy markets away from emerging market crises and reduced the pressure to find real consensus on difficult issues. Worries about too much demand for IMF lending gave way to concerns that the IMF might earn too little from its lending to cover its expenses. Concerns that the IMF was too reluctant to criticize emerging economies that clung too tenaciously to exchange rate pegs in the face of current account deficits gave way to concerns that the IMF was too reluctant to criticize emerging economies that intervened heavily to maintain undervalued exchange rates. Concerns that emerging economies had too few reserves gave way to concerns that they held too many   and were adding even more every year.

By 2005, private capital flows to emerging economies matched their pre Asian crisis peak. Eastern Europe, in part because of its institutional tie to the rest of Europe, relied on these capital inflows to cover significant current account deficits. But most other emerging economies used this surge in private capital flows, along with the windfall from rising commodity prices, to add to their reserves. In aggregate, the emerging world ran a large current account surplus, financing deficits in many of the world's advanced economies. This system, which has been labeled a new Bretton Woods system, represented a real change in the international financial architecture, even if it did not emerge from the debate that occupied policymakers in the 1990s, let alone a new Bretton Woods conference.

The debate over the architecture for managing financial crises in emerging economies has been superseded by a different architecture debate focused on finding new ways to facilitate macroeconomic policy coordination between the world's new creditor countries and its major debtor countries to help reduce global imbalances. This debate covers a very different set of topics than the previous debate from the optimal use of emerging economies' vast reserves to strengthening the IMF's surveillance of surplus countries in the emerging world and deficit

countries in the G7. One topic, though, has stayed constant: the continuing need to adapt the international financial world's governing institutions, both formal and informal, to better reflect the growing role emerging economies play in the international financial system.

Also see bail-ins; bailouts; banking crisis; Bretton Woods system; capital flows to developing countries; currency crisis; exchange rate regimes; financial crisis; global imbalances; hedge funds; hot money and sudden stops; inflation targeting; International Monetary Fund (IMF); International Monetary Fund conditionality; International Monetary Fund surveillance; international policy coordination; international reserves; Latin American debt crisis; lender of last resort; original sin; twin deficits

FURTHER READING

Group of Seven. 1999. *Strengthening the International Financial Architecture: Report of G 7 Finance Ministers to the Cologne Economic Summit.* Cologne, Germany: G7. A comprehensive statement of the G7's agreed approach to the architecture reform.

International Financial Institutions Advisory Commission (IFIAC). 2000. *Report.* Washington, DC: United States Congress. Downloadable from http://www.house.gov/jec/imf/meltzer.htm. More commonly called the Report of the Meltzer Commission. It proposed that the IMF lend for very short terms only to countries that prequalified for support.

Kenen, Peter B. 2001. *The International Financial Architecture: What's New? What's Missing?* Washington, DC: Institute for International Economics. A retrospective assessment of the initial round of reforms from an influential academic voice.

King, Mervyn. 2006. "Reform of the International Monetary Fund." (February). A key speech that helped to define the current "architecture" debate.

Krueger, Anne. 2001a. "International Financial Architecture for 2002: A New Approach to Sovereign Debt Restructuring." Speech at the National Economists' Club Annual Members' Dinner, American Enterprise Institute, November 26. Washington, DC. Downloadable from http://www.imf.org/external/np/speeches/2001/112601.htm. The case for a new international bankruptcy regime.

Rubin, Robert E., and Jacob Weisberg. 2003. *In an Uncertain World: Tough Choices from Wall Street to Washington.* New York: Random House. Rubin's thinking on the international financial architecture.

Task Force on the Future International Financial Architecture (CFR Task Force). 1999. *Safeguarding Prosperity in a Global Financial System.* New York: Council on Foreign Relations. An agenda for reform that both parallels and differs from the reforms the G7 embraced at Cologne.

Truman, Edwin. 2006. *A Strategy for IMF Reform.* Washington, DC: Institute for International Economics. A comprehensive review of current proposals for IMF reform.

BRAD SETSER

■ international financial centers

International financial centers (IFCs) are global hubs for banking, insurance, capital, money, and foreign exchange transactions. A large share of global trade and management of financial wealth and debt instruments is conducted in these centers. By virtue of the volume of transactions and the depth of their markets, they are leaders in setting the price of financial instruments. Such instruments are used to channel (intermediate) capital between savers and borrowers and to transfer financial risk.

The predominant IFCs are London and New York. Other key centers include Tokyo and Frankfurt. A number of financial centers are important in a regional context, including Singapore and Hong Kong. Offshore financial centers are important domiciles for international financial transactions with the pricing and trading of the financial instruments keyed off the markets in the major IFCs. The term *international financial center* is also used as a signal by some centers that a preferential tax and/or flexible regulatory regime will apply to financial entities/transactions domiciled there. Competition among the financial centers is intense, and successful centers offer comparative advantage for market participants.

Growth and Evolution of IFCs The major IFCs have evolved in response to underlying economic circumstances. The initial importance of both London and New York can be traced to their geographical and historical locations as centers of trade and commerce and to the use, respectively, of British pounds and U.S. dollars for the denomination of international trade and investment transactions. Through a series of financial innovations, London has remained a dominant IFC notwithstanding the post World War II decline in the importance of sterling and the UK economy. Since the 1960s, for example, London has been the principal host of an international market in money, credit, and debt instruments denominated in U.S. dollars (the so-called Eurodollar market). Other innovations include deregulation of its stock market in the mid-1980s, elimination of exchange controls, and a market-friendly approach to regulation. The volume of international transactions in London is now several times that of domestic transactions. A similar balance of transactions can be found in centers such as Hong Kong, Singapore, Luxembourg, Ireland, and Switzerland, for example. By contrast, domestic transactions still dominate in centers such as New York and Tokyo, reflecting the size of their domestic economies.

Several factors have promoted the development of IFCs. The growth of international trade and investment in the period following World War II was accompanied by a rapidly expanding volume of international capital flows. The global increase in capital movements has responded to and helped finance the emergence of substantial imbalances between countries with surpluses and those with deficits in their balance of payments. A significant part of these international capital flows has been intermediated by the institutions and markets in the IFCs as they adapted their financial institutions and instruments to the global financing needs. Large deficits in the U.S. current accounts have been matched by increases in purchases of U.S. dollar assets, primarily through the major IFCs.

During the 1970s, a succession of increases in oil prices resulted in significant balance of payment surpluses among oil-producing countries. These surpluses were "recycled" through the IFCs, initially largely through the placement of surplus funds in bank deposits in those centers. At the same time, a number of advanced, developing, and emerging market economies looked to the international financial markets to finance their foreign-exchange reserve positions or their programs of economic growth and development. The facilities and instruments to mobilize the capital transfers, such as syndicated loans and international bonds, were provided by institutions based in the IFCs. Large corporations also tapped the international markets to finance investment.

A theoretical explanation of why certain financial centers have become dominant can be found in the literature on central place theory, which is used to explain why cities tend to be of unequal size. The underlying idea is that there is a trade-off between economies of scale that lead to concentration of activities, spurring centers of activity to grow, and the costs of accessing these centers of activity, which places a limit on the size of the center and allows other centers to grow. In the case of financial transactions, potentially significant benefits of concentration include the pooling of expertise and the bringing together of a large number of buyers and sellers of financial instruments. The ability to trade with a financial center can also be limited, for example, by the time zone where the financial center is located. The central-place theory has limitations when it comes to financial centers, which are in intense competition. No center can remain dominant as an IFC relying solely on its location. In the first decade of the 21th century New York started falling behind, losing traditional businesses to London, while new centers sprang up, as in Dubai, for example, which captured business from Europe, Asia, and the Middle East.

IFCs are characterized by mature financial markets that are deep (have a capacity to handle a range of transactions of different tenor and in large volumes), liquid (have a capacity to trade in and out of positions with reasonable certainty about the prices at which the trades will take place), and relatively efficient

(have pricing that reflects current market information). The legal foundation for financial contracts is well developed, providing reasonable legal certainty in the execution of the transactions, combined with sufficient flexibility to accommodate rapid innovation in financial instruments. The technical and institutional infrastructures for the financial markets are advanced and include a highly skilled and specialized labor force, ease of communications, efficient and low-cost trading platforms, and payments systems that help to mitigate financial systems risks. These systems are supported by regulatory structures and supervisory systems directed to customer protection, the prevention/mitigation of market and institution failures, and the prevention of financial abuse. The concentration of financial activity in the major centers places a significant responsibility on the host countries to ensure that the financial markets and institutions meet the highest regulatory standards, not least to protect their reputations, and regulators and central banks have proved quite adept at intervening at critical moments to maintain the stability and reputation of their centers.

The importance of the major business lines varies among the different financial centers, and they have also been evolving rapidly under intense competition and financial innovation. These business lines include (1) international banking (intermediating between international borrowers and depositors), investment banking (e.g., underwriting of equities and fixed-income investments, underwriting and brokering securitized streams of payments obligations [e.g., from credit cards, mortgages] and advisory services in mergers and acquisitions), and private banking (investment and advisory services for private wealth management); (2) foreign-exchange trading in spot, forward, and swap markets covering all the major currencies as well as nondeliverable forward markets in less convertible currencies; (3) insurance activities of various forms (e.g., life, marine, property, and casualty) and reinsurance; (4) securities dealing in equities, primarily through organized exchanges, and bonds in domestic and international currencies; (5) fund and asset management and administration such as mutual funds, pension funds, and hedge funds; (6) derivatives trading (e.g., swaps, options, interest rate equities, commodities, and credit contracts) both in over-the-counter markets and on organized exchanges; and (7) commodities trading (e.g., precious and other metals, agricultural and energy products) in spot, futures, and options transactions.

Consequences of Growth of IFCs IFCs are important both to the economies where they are located and internationally. They have direct importance to the local economies as sources of employment, contributors to foreign earnings and the balance of payments, and to gross domestic product. Internationally their importance derives from the provision of a range of financial services more efficiently than would be possible in local financial markets, reflecting the volume of transactions and the sophistication of the supporting infrastructure. The major banking and financial groups, which have a large international presence, are also headquartered in these centers. The management of a large share of global financial activity is therefore directed through these centers. The institutions in the IFCs have traditionally been at the forefront of innovations in financial instruments and techniques, adapting rapidly to changes in the global financial environment.

The global consequences of IFCs have not been without controversy. The growth of the Eurocurrency market was associated with concerns that it would undermine monetary policy and result in an uncontrolled expansion in global credit. These concerns abated as the authorities moved from monetary control that relied on direct instruments (such as credit ceiling and high, non-interest-bearing reserve requirements) which were not applied to transactions through the euro markets to indirect monetary instruments that relied on the market's adjustment of interest rates. The latter has a global impact, including in the euro markets. From time to time there are concerns that competition in the IFCs leads to an underpricing of risk, threatening the stability of financial institutions, and the global financial system. Credit spreads have tended to follow patterns of decline associated with financial innovation, only to

face sudden reversals when the assumptions and instruments on which the declines were based are tested by worsening economic conditions. Examples include the decline in spreads in the 1970s and 1980s on credits to developing countries with the growth of the market in syndicated credits that preceded the first international debt crisis; and the expansion of securitization to a broader range of debt payment flows, including subprime mortgages that preceded a sharp loss of confidence in the international credit markets in 2007. Rapid innovation in financial centers has posed, and continues to pose, a major challenge for financial regulators as they seek to understand and mitigate potential and emerging risks. As the underlying transactions have become less transparent, the rating agencies have grown to play an increasingly important role in the pricing of credits and the functioning of markets. The IFCs have provided global benefits, but they are not without risk, and history has shown that the markets and their regulators will be tested.

See also Eurocurrencies; offshore financial centers; petrodollars, recycling of; trade in services

FURTHER READING

Johnston, R. B. 1983. *The Economics of the Euro Market: History, Theory, and Policy*. London: Macmillan Press. Surveys the reasons for the development of IFCs, their role in global finance, and the policy debates on regulation of these centers.

R. BARRY JOHNSTON

■ international income convergence

Observers of the world economy often ask whether poorer countries are catching up with the rich countries. Put differently, are average living standards converging? This fundamental question has been much studied, especially since the development of new models and data in the mid-1980s. The answers help to shed light on the effects of globalization and development policies and provide a way of thinking about long-run predictions for the world economy. In principle, they also help to discriminate between models of the growth process. For all these reasons, the convergence question is central to the study of growth and development and also plays a role in international economics.

In popular commentary, it is common to hear claims such as "the rich countries are getting richer and the poor are getting poorer." It is well known that, after the mid-19th century, income levels diverged as some countries grew faster than others, but whether the poor countries became even poorer is open to debate. For more recent decades, the pattern is more complicated. Some poor countries have grown rapidly, notably many of those in East and Southeast Asia. Others have not, and some of the worst performers, predominantly in sub-Saharan Africa, have seen their gross domestic product (GDP) per capita decline even in absolute terms.

The diversity of experience implies that detailed studies of convergence may have an informative role to play. These studies are usually based on comparisons in relative terms, using ratios (rather than absolute gaps) to assess inequality in living standards or labor productivity at different points in time. Measurement should allow for differences in price levels across countries, and this can be achieved using version 6.2 of the Penn World Table developed by Heston, Summers, and Aten (2006). Using the standard deviation of the logarithm of output per capita to measure inequality, and excluding countries with populations below 250,000 in 1960, reveals a modest but steady increase in dispersion between 1960 and 2003 for a sample of 93 countries. The standard deviation rose by around 25 percent over the period as a whole.

The pattern is more complicated if we concentrate on the developed countries. For the countries above the 80th percentile of output per capita in 1960, there is some convergence until the late 1970s, but divergence later. There is clearer evidence for convergence if we restrict the sample to the Western European economies. In summary, there is no general tendency for convergence, unless we consider subgroups of similar economies, such as those of Western Europe. These findings will be one theme of what follows.

Convergence and the Solow Model Our understanding of convergence and divergence needs to be guided by theoretical models. Much of the literature, especially in its empirical guise, starts from the one-sector neoclassical growth model of Solow (1956). An aggregate production function is used to relate a country's output to a set of inputs, which would typically include capital equipment and labor. In this view of the world, international differences in labor productivity arise from differences in inputs per worker, or from differences in how efficiently inputs are combined. Convergence then requires stocks of inputs per worker, or efficiency levels, to move closer together. If GDP per capita is the yardstick, rather than output per worker hour, we should also consider changes in labor force participation rates and average working hours.

Models of this kind embody a view of living standards that is very different from the accounts found in popular commentary. In those accounts, development is often seen as a competitive process, in which wealth in one country is said to require poverty elsewhere. In contrast, in many theoretical models, productivity is determined independently of the productivity levels of other countries. If countries abroad are growing rapidly, this does not undermine domestic living standards.

In exploring the other predictions of growth models, the literature has distinguished between two different ways in which countries might be said to converge. *Absolute convergence* is the form that most commentators have in mind, namely whether inequality across countries in average living standards is tending to diminish. As we have seen, the overall tendency in recent decades has been one of gradual divergence.

We can learn more about this using a second idea, that of *conditional convergence*. The central hypothesis is that convergence is only to be expected within groups of economies that share similar structural characteristics. Within such groups, we should expect the countries that are initially poorer to grow more quickly, but this tendency is still compatible with divergence overall, given that there may be important differences across groups.

For this idea to be more than a truism, we need to be specific about the characteristics that are similar within groups. Consider a simple version of the Solow model, in which technical efficiency grows at the same rate in all countries. In this case, countries will converge to parallel equilibrium growth paths. Along these paths, capital and output will grow at the same rate, which is the equilibrium outcome, or steady-state, implied by the model. We can now consider what the model predicts about relative levels of output per worker in the long-run equilibrium   in other words, the extent of dispersion of the equilibrium paths. The highest levels of output per worker will be associated with countries that have high rates of investment, high levels of efficiency, and low rates of population growth. If two countries are alike in these various respects, the country that is initially poorer will grow relatively quickly, and living standards in the two countries will converge. But this convergence is only conditional, because it relies on the two countries looking alike in terms of the variables that influence the level of the steady-state growth path. This can explain why we might observe absolute convergence only within subgroups of countries, such as those of Western Europe, that are likely to share similar steady-state paths.

Even when conditional convergence is a good description of the data, whether or not countries converge in the absolute sense will depend on a range of factors. These include not only the dispersion of the equilibrium growth paths, but also the starting point of each country relative to its own equilibrium path. A further complication is that variables such as investment rates or population growth rates may be changing over time. In this case, the equilibrium paths of the different countries will move gradually apart or closer together, and this too will influence the observed extent of convergence.

The same basic analysis can easily be extended in other directions. For example, the quality of a country's institutions may be an important determinant of the level of the steady-state growth path. Other candidates include the education and skills of a country's workforce and the extent of government-provided infrastructure. The explicit specification

of steady-state determinants is central to the Solow model's empirical content and the testing of its predictions.

Convergence predictions become more complicated in models where there is more than one equilibrium outcome. For example, the Solow model can be adapted to make the population growth rate a declining function of the level of income per capita. In that case, there may be one equilibrium with high living standards and slow population growth, and another equilibrium with low living standards and faster population growth. Although models with more than one equilibrium outcome are inherently difficult to test, they support the idea that income dispersion is not the only relevant issue. We might also be interested in whether there is any mobility within the distribution of living standards and whether the distribution changes shape over time. The work of Quah (for example, 1996) represents the major contribution along these lines.

There are several other limitations to the analysis sketched above. One is that we have assumed that the rate of technical progress is the same across countries, so that all countries grow at the same rate in the long run. This assumption is unrealistic, but less so than it first appears. Imagine that, other things equal, countries with a low initial level of efficiency are more likely to show rapid improvements in efficiency, perhaps because there is greater scope for easily adopting existing technologies from abroad. In that case, there may still be an equilibrium in which all countries grow at the same rate in the long run. In this long-run equilibrium, differences in the speed of adoption will be reflected in equilibrium "technology gaps" between leading and following countries, so that countries differ in relative income levels rather than long-run growth rates.

Convergence among Interdependent Economies The simplicity of the Solow model can be useful, not least in developing some intuition about where and when convergence should be expected. It is clear that absolute convergence is far from inevitable and that, even in a simple model, a wide range of outcomes is possible. The model is also a useful corrective to the view that economic development is

essentially a competitive process and helps to explain why that view is less popular with economists than with noneconomists. Nevertheless, it is obvious that much of interest has been omitted. In practice, countries exchange goods and services; capital and labor sometimes move between countries; and ideas and technologies may spread beyond national borders in complex ways. These interactions are worth considering in more detail.

Many observers see the international trade of goods and services as central to the convergence process. It is true that, in principle, trade can lead to convergence in factor prices such as real wages and the returns to capital. In this case, differences in living standards will arise only when inputs per capita differ. If capital and labor are the only inputs used in production, factor price equalization implies that differences in living standards are driven entirely by differences in capital-labor ratios.

There are two problems with this story. First, it is false even at the level of casual observation, because real wages are readily seen to vary widely across countries. Second, the theoretical conditions under which factor price equalization occurs are extremely strict. Typically, production technologies are required to be the same across countries, but recent empirical work usually concludes that levels of technical efficiency vary widely. In this case, trade may still equalize factor prices, but only once we adjust them for international differences in the productivity of the different factors. The required adjustment makes this more general prediction harder to test.

Even in the absence of factor price equalization, trade in goods and services may still contribute to the convergence process. One view is that trade helps to encourage the transfer of technology across countries. These "dynamic" gains from trade are less well understood than the static gains of conventional trade theory. Their possible importance is highlighted by countries that have grown rapidly while becoming more integrated with the world economy. Separating this process of integration and growth into cause and effect is not straightforward, and empirical researchers seem to be divided on the

extent to which trade flows promote technology diffusion and convergence. It is also worth noting that introducing dynamic considerations can overturn some of the predictions of textbook models of trade. For example, firms and countries may become more productive as they gain experience in producing a particular good or service. If this process of "learning-by-doing" has an important role in productivity improvement, then sometimes trade can even lead to divergence (see Helpman 2004 for more discussion).

Since the relationship between trade and convergence is potentially ambiguous, empirical evidence becomes especially important. One of the best-known studies is that by Sachs and Warner (1995). They classify most of the world's economies as either closed or open to trade. Over the period 1970 89, they find evidence for convergence within the group of open economies, but not within the group of closed economies. Open economies are also found to grow more quickly on average. The results are hardly conclusive, however. A high proportion of the open economies in the sample are members of the Organisation for Economic Co-operation and Development, and perhaps these countries converged because they are alike in other ways. Moreover, when Sachs and Warner distinguish between closed and open economies, their criteria can be interpreted as reflecting several broader dimensions of economic policy, and not simply trade policy.

Another approach asks whether specific episodes of trade liberalization have been followed by convergence among the liberalizing countries. Slaughter (1997) provides a summary and critique of some of this work. An ever-present constraint is the small number of countries available to empirical researchers, which makes it difficult to establish whether trade liberalization promotes convergence in practice.

If the effects of trade on convergence are uncertain, those of international capital flows are even more so. As is well known, these flows increased over the 1980s and 1990s, as many countries dismantled capital controls. The move to open capital accounts means that domestic investment is no longer constrained by domestic saving, and capital will tend to move to wherever the return is highest. Countries that are initially poor, with low saving rates, could be among those to benefit from this process. In textbook models, the marginal product of capital will be especially high in these countries, and capital will flow into them until the returns are driven down to the level available elsewhere in the world. Output and wages have the potential to increase rapidly. Although profits associated with the new investment will be returned to the foreign owners of capital, national income will also increase over time.

This mechanism only works, however, if foreign investors are not deterred by high risks. Moreover, countries that receive large-scale capital flows have the potential to be destabilized by, for example, changes in investor sentiment. This argument gained increased prominence with the East Asian financial crisis of 1997 98. To the extent that capital flows are destabilizing, their effects on convergence must be regarded as ambiguous.

Such problems may be less serious for long-term capital flows, notably foreign direct investment (FDI). It is often suggested that FDI encourages the international diffusion of technology, promoting convergence. The effects of FDI on productivity in recipient countries, and the extent of spillover benefits for domestic firms, have been a focus of recent empirical research.

This introduction only skims the surface of the theory and evidence on convergence. In this respect, one final point is worth making. In popular commentary on issues such as globalization, or development policies, the outcomes are sometimes assessed primarily in terms of inequality between countries. This is not without its associated dangers. In principle, it is conceivable for some events or policies to raise incomes everywhere, while leading to overall divergence in living standards, if some countries benefit more than others. This suggests that the analysis of convergence development in relative terms should often be supplemented with consideration of the level of development in absolute terms. It is the latter that is likely to have more influence on the everyday lives of the world's poor.

See also capital accumulation in open economies; capital mobility; foreign direct investment and international technology transfer; growth in open economies, neoclassical models; trade and economic development, international

FURTHER READING

Dowrick, Steve, and J. Bradford De Long. 2003. "Globa lization and Convergence." In *Globalization in Historical Perspective*, edited by Michael D. Bordo, Alan M. Tay lor, and Jeffrey G. Williamson. Chicago: NBER and University of Chicago Press, 191 220. Takes a wide, historical view of convergence.

Galor, Oded. 1996. "Convergence? Inferences from The oretical Models." *Economic Journal* 106 (437): 1056 69. Examines the convergence implications of models with more than one equilibrium outcome.

Helpman, Elhanan. 2004. *The Mystery of Economic Growth*. This book, especially chapter 5, is one of the best in troductions to these issues.

Heston, Alan, Robert Summers, and Bettina Aten. 2006. "Penn World Table Version 6.2." Data set available at the Center for International Comparisons of Produc tion, Income, and Prices, at the University of Pennsyl vania (September).

Jones, Charles I. 2002. *An Introduction to Economic Growth*. 2d ed. New York: Norton. An excellent introduction to a range of growth models.

Mankiw, N. Gregory, David Romer, and David Weil. 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107 (2): 407 37. A classic paper in growth economics, partly because it gives new empirical content to the Solow model and clarifies the role of conditional convergence.

Obstfeld, Maurice, and Kenneth Rogoff. 1996. *Foundations of International Macroeconomics*. Cambridge, MA: MIT Press. Chapter 7 provides an illuminating discussion of many aspects of convergence, with a focus on open economies.

Quah, Danny T. 1996. "Twin Peaks: Growth and Con vergence in Models of Distribution Dynamics." *Eco nomic Journal* 106 (437): 1045 55. One of several classic studies by Quah on the world's distribution of income per capita across countries.

Sachs, Jeffrey D., and Andrew W. Warner. 1995. "Eco nomic Reform and the Process of Global Integration."
Brookings Papers on Economic Activity 1: 1 118. An in fluential but controversial study of the links between national policies and convergence.

Sala i Martin, Xavier. 2006. "The World Distribution of Income: Falling Poverty and . . . Convergence, Period." *Quarterly Journal of Economics* 121 (2): 351 97. A study of income inequality and convergence across the world's individuals, rather than across countries.

Slaughter, Matthew J. 1997. "Per Capita Income Con vergence and the Role of International Trade." *American Economic Review* 87 (2): 194 99. A useful overview of some of the convergence mechanisms that arise when countries trade with one another.

Solow, Robert M. 1956. "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics* 70: 65 94. Introduced Solow's classic model of national economic growth.

JONATHAN R. W. TEMPLE

■ **international institutional transfer**

There is by now a substantial literature on the im portance of institutions to economic development. Institutions are formal or informal rules that con strain individual choice and reduce transaction costs, thereby facilitating collective action (North 1990). Governments that can credibly commit to the en forcement of property rights and a rule of law that can adjudicate conflicting property rights claims are seen as particularly important in facilitating long term investment and therefore growth (de Soto 1989). In addition, political institutions that miti gate social conflict and provide public goods are also critical. There is substantial empirical evidence that institutions are strongly correlated with growth (see Knack and Keefer 1995; Easterly and Levine 2002); there is a debate, however, as to whether institutions are endogenous or exogenous to the economic system (see Sachs 2005).

Colonialism One of the most common exoge nous sources of institutions are foreign actors, in the form of colonial powers, occupation authorities, industrialized great powers, multilateral organiza tions such as the World Bank or the International

Monetary Fund, or other international donors. There is considerably less empirical understanding of possibilities for institutional transfer across international boundaries than there is for the endogenous creation of institutions within a single society.

One important source of information concerning institutional transfer comes from the history of colonialism, where mostly European great powers exercised formal sovereignty over territories in what is now the developing world. There is, unfortunately, a relatively small comparative literature that systematically analyzes the long-term institutional impact of different colonial regimes. This is, of course, a morally fraught subject, since most colonial relationships were exploitative and destructive of indigenous institutions and cultures. It is clear that there is a huge variance in colonial experiences, even within individual empires. The British, for example, left relatively durable institutions in the Indian subcontinent and in their settler colonies in southern Africa, but in other places such as West Africa they relied largely on local elites and left behind considerably weaker institutions. The Japanese were relatively more successful in creating durable economic institutions in their Korean, Taiwanese, and Manchurian colonies (Woo 1991) than the Americans were in the Philippines. Acemoglu, Robinson, and Johnson (2001) suggest that European colonialists left strong institutions in territories where environmental conditions (particularly disease burdens) allowed them to settle, such as North America and South Africa. In other places, such as Latin America and West Africa, they created extractive regimes that were not conducive to securing property rights or rule of law. In a similar vein, Engerman and Sokoloff (2002) have suggested that initial factor endowments the nature of cash crops, mineral resources, and labor predisposed institutions in North and South America and the Caribbean to greater or lesser degrees of inequality, differences that have persisted over time.

The Governance Agenda in International Development Decolonization and the emergence of the developing world in the 1950s, 1960s, and 1970s created new conditions for institutional transfer. In the late 20th and early 21st centuries, foreign powers usually did not have sovereign authority over the territories to which they sought to transfer institutions but had to use indirect methods such as foreign assistance, technical advice, and conditionality in lending for structural adjustment and other purposes. There continued to be instances of direct rule by international actors, however, under postconflict conditions in places such as Haiti, Somalia, Bosnia, Kosovo, Timor-Leste, Afghanistan, and Iraq (see Chesterman 2004). Over the years, the international donor community has been increasingly concerned with improving the governance and institutional performance of developing countries, to the point where a substantial proportion of lending by multilateral institutions such as the World Bank and the Inter-American Development Bank are linked to public sector reform (World Bank 1997, 2000, 2002).

It is difficult to assess the overall impact of these external interventions designed to strengthen institutions and governance. In certain regions of the world such as Latin America, there has been a clear improvement in public sector performance, particularly in the area of macroeconomic policy management (Lora 2007). On the other hand, there has been much less progress in strengthening judicial systems, fighting corruption, or improving the performance of social sector institutions in education and health. In either case, it is not clear the extent to which existing institutional reform has been the product of external advice, help, and pressure, or internally driven reform processes. The World Bank's program to improve financial management in Highly Indebted Poor Countries has yielded very ambiguous results (Levy and Kpundeh, 2004). In Timor-Leste, the United Nations Transitional Authority for East Timor undertook the direct rule of a new nation between 1999 and 2002; while this was initially regarded as a big success for international state-building, Timor-Leste's army and police collapsed in internal fighting in 2006. The American experience in trying to create liberal democracies in Afghanistan and Iraq has been, to say the least, disappointing.

Theorizing Governance Recent efforts to improve governance have been conceptualized under a principal-agent framework. Governments are understood to be hierarchies in which principals (in a democracy, the sovereign people) direct agents to perform certain activities. The agents, however, have their own interests, which are not always aligned with the interests of the principals. Improving governance is therefore understood to be a matter of aligning agent incentives with those of their principals, largely by creating systems of monitoring and accountability.

Public sector services can be characterized along two dimensions: transaction volume and specificity. *Transaction volume* refers to the number of decisions or other outputs that a particular public sector agency has to make; they can range from small, in the case of a central bank or finance ministry, to large, in the case of a judicial system or public education system. *Specificity* refers to the degree to which the quality of the output of a public sector agency can be externally monitored. Macroeconomic policy institutions such as central banks and finance ministries have much lower transaction volume and higher specificity than institutions providing public education or legal services; their performance can therefore be much more easily monitored and brought under a principal-agent framework. This explains why most improvement in institutional performance in recent years has come in low-transaction-volume, high-specificity parts of the public sector.

This framework indicates why legal systems protecting property rights are particularly hard to transfer. Legal systems involve high transaction volumes, and their output is very difficult to measure. Property rights can be and often are selectively enforced, leading to rent-distribution systems with highly variable effects on economic growth (Khan and Jomo 2000).

Why Institutions Are Difficult to Transfer More broadly, there are four reasons why institutions are generically difficult to transfer across national boundaries. The first is that there are often large problems of fit between desired institutions and the nature of the underlying society. The models for well-functioning states in the minds of Western donors are often idealized versions of their own political institutions, which in turn are based on some vision of a Weberian rational-bureaucratic state. These embed certain clear value choices and moral perspectives: that exchange should be impersonal, that obligations to the society as a whole should trump narrower communal attachments, that hiring and advancement should not be based on kinship but on merit, and so on. These values do not come naturally to any human society and emerged in the West and in parts of East Asia only after a prolonged historical process. It is therefore not surprising that there is a poor fit in virtually all developing countries between the existing normative order and social structure and the kinds of institutions that donors prefer. Many donors believe that the burden falls on the developing country to adjust to this idealized model. The adjustment, however, must be mutual. There are no "optimal" forms of institutions; among developed liberal democracies, there is a wide variance in the ways institutions are implemented that make them "fit" their societies better.

The second obstacle to institutional reform concerns the problem of transmission mechanisms and ownership. The objectives of service delivery and capacity-building are often at odds; because public sector services are weak or nonexistent, donors are often tempted to provide them directly, which then weakens the capacity of the reforming country's government to provide them over the long term. Most successful state-building projects, such as those of the East Asian fast developers, have been driven by domestic elites; the latter may have gotten help in the form of resources or technical assistance from outside allies or donors, but themselves created demand for reform. In the absence of such internal demand, donors are often tempted to stimulate it artificially, either through conditionality in lending to mechanisms like the European Union's accession criteria. America's Millennium Challenge Account seeks to provide similar incentives for governance reform. Alternatively, donors can seek to promote independent civil society and independent media, which, it is hoped, will unlock a latent demand for change by

putting pressure on corrupt or incompetent governments. Although breakthroughs such as Ukraine's Orange Revolution occasionally occur, externally stimulated demand for reform is seldom sufficient to produce decisive results.

The third obstacle is that state-building is at base a political process and not a technical one, one that involves access to power and resources and is therefore highly contested. Most dysfunctional public bureaucracies can trace their poor performance directly to the intervention of politicians who want to use the administrative machinery for their own purposes or prevent it from interfering in their activities. This, of course, presents a major obstacle for external donors, whose influence on local politics is either prohibited by statute, as in the case of the World Bank, or else is limited by past historical colonial relationships. Administrative reform pursued as a technical assistance project can get only so far if political will is lacking, and donors often do not have the patience to wait for the emergence of the right political conditions.

The fourth obstacle has to do with the need for nation-building in addition to state-building. It is common for those involved in postconflict reconstruction to concentrate on improving the capacity of public sector institutions and to shy away from attempts to help formulate or strengthen national identity, common culture, or shared values within the client country's society. Yet the latter is critical for the success of the former in the long run. If bureaucrats in a public agency, for example, do not identify the public good with that of the nation as a whole, but rather see their purpose as helping one particular ethnic or kin group advance relative to others, then that agency's purposes will be corrupted from the start. Outsiders usually have limited means of fostering national identity. Although it has been done in the past (for example, by the British in India), building a national identity is a very slow process and does not suit the timetables of most donors.

Future Directions The difficulty of transferring institutions has led to suggestions that institutions need to be, in effect, outsourced. Krasner (2004) has suggested the need for a system of "shared sovereignty" in which certain state functions are performed by international organizations or other external actors. Other writers have suggested that it is often well-meaning interventions on the part of the international community that create state weakness in the first place, and that states may be better off being left alone to generate institutions endogenously (Weinstein 2005). It is clear, however, that the realities of globalization mean that institutional development in many parts of the world will be inescapably linked to external actors, ideas, resources, and influences. To an increasing degree, it is impossible to separate the exogenous and endogenous, internal and external, sources of growth and decay of institutions.

See also aid, international; development

FURTHER READING

Acemoglu, Daron, James A. Robinson, and Simon Johnson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91 (5): 1369 401. Widely cited article attributes institutional transfer to disease burdens, which determine the likelihood of Europeans settling in a particular colony.

Chesterman, Simon. 2004. *You the People: The United Nations, Transitional Administration, and State Building*. Oxford: Oxford University Press. A comprehensive study of post Cold War, postconflict interventions by the UN.

De Soto, Hernando. 1989. *The Other Path: The Invisible Revolution in the Third World*. New York: Harper and Row. A classic account of informality in developing countries as a consequence of weak formal institutions.

Easterly, William R., and Ross Levine. 2002. *Tropics, Germs, and Crops: How Endowments Influence Economic Development*. NBER Working Paper No. 9106. Cambridge, MA: National Bureau of Economic Research. Analysis confirms the importance of institutions relative to factor endowments and geography.

Engerman, Stanley L., and Kenneth L. Sokoloff. 2002. "Factor Endowments, Inequality, and Paths of Development among New World Economies." *Economia* 3 (1): 41 101. Authors suggest that factor endowments

determined the nature of institutions and inequality when comparing Latin and North America.

Fukuyama, Francis. 2004. *State Building: Governance and World Order in the 21st Century*. Ithaca, NY: Cornell University Press. A general account of the role of states in development and of the limits to our knowledge about institutional transfer.

Khan, Mushtaq H., and Kwame Sundaram Jomo, eds. 2000. *Rents, Rent Seeking, and Economic Development: Theory and Evidence in Asia*. Cambridge: Cambridge University Press. Shows that rent seeking is not uni formly harmful to growth and suggest a typology of rent seeking regimes.

Knack, Stephen, and Philip Keefer. 1995. "Institutions and Economic Performance: Cross Country Tests Using Alternative Measures." *Economics and Politics* 7: 207 27. An early econometric study linking institutions to growth.

Krasner, Stephen D. 2004. "Sharing Sovereignty: New In stitutions for Collapsed and Failing States." *International Security* 29 (2): 85 120. Suggests that, given the diffi culty of institutional transfer, permanent shared sover eignty with international actors may be the only possible way of providing good governance for failed states.

Levy, Brian, and Sahr Kpundeh, eds. 2004. *Building State Capacity in Africa: New Approaches, Emerging Lessons*. Washington, DC: World Bank. Surveys the success of the World Bank's public sector reform efforts in sub Saharan Africa.

Lora, Eduardo, ed. 2007. *The State of State Reform in Latin America*. Stanford, CA: Stanford University Press/IDB. This collection of studies of institutional reform in Latin America since the debt crisis shows the glass half full, with importance governance gains in some areas and stagnation or regress in others.

North, Douglass C. 1990. *Institutions, Institutional Change, and Economic Performance*. New York: Cambridge University Press. Provides the classic framework for understanding institutions in terms of transaction costs.

Sachs, Jeffrey. 2005. *The End of Poverty: Economic Possibi lities for Our Time*. New York: Penguin Press. Sachs returns to the development policies of the 1950s with his call for an injection of massive donor resources to over come the "poverty trap" in which he argues much of Africa is stuck.

Weinstein, Jeremy. 2005. *Autonomous Recovery*. Wa shington: Center for Global Development. A passionate but not always persuasive plea for greater resources as a means of spurring development.

Woo, Jung En. 1991. *Race to the Swift: State and Finance in Korean Industrialization*. New York: Columbia Univer sity Press. Indicates the importance of Japanese colonial institutions to South Korea's postwar development.

World Bank. 1997. *The State in a Changing World*. Oxford: Oxford University Press. This report marked an im portant shift away from the Washington consensus in the Bank's internal thinking, giving new emphasis to the importance of the state for development.

———. 2000. *Reforming Public Institutions and Strength ening Governance*. Washington, DC: World Bank. This World Development Report marked an important turning point in the World Bank's thinking about the centrality of institutions and governance to development.

———. 2002. *Building Institutions for Markets*: *World Development Report 2002*. Washington, DC: Oxford University Press. A follow up to the 1997 World De velopment Report that provides further conceptual and empirical grounding for the Bank's focus on institutional reform.

FRANCIS FUKUYAMA

■ international investment agreements

Investment provisions in bilateral investment treaties and reciprocal preferential trade agreements (PTAs) are intended to promote investment flows by grant ing investors greater predictability of the policies regulating foreign investment. These agreements typically provide for transparency, national treat ment, nondiscrimination among foreign investors, and guarantees against expropriation. Some agree ments may include disciplines preventing trade-re lated investment measures (TRIMS), such as local content requirements and local hiring requirements. Finally, nearly all provide some sort of dispute set tlement provisions.

The Rise of Bilateral Investment Treaties (BITs)
BITs are the primary vehicle for international co operation on the regulation of investment flows.

They customarily provide a definition of investment coverage, include rules of origin to determine the nationality of investors, and specify the treatment of inward investment and investors, either once established and/or preestablishment. Most BITs do not deal with market access restrictions per se — such as restrictions on sectoral entry or equity ownership limitations — although some do impose disciplines on performance requirements and similar TRIMs. They generally guarantee most favored nation (MFN) and national treatment (banning discrimination among investors from signatory countries and guaranteeing treatment of foreign investors on the same footing as domestic investors, respectively). Often the treatment that is guaranteed foreign investors is actually better than that accorded to domestic investors, for example, in terms of access to foreign exchange or ability to transfer capital outside of the country, or in terms of investor protection/rights.

A major function of BITs is to provide investor protections such as against expropriation, stipulating that there be due process, transparency and compensation for any expropriation of foreign investments. They also provide for a dispute resolution mechanism to enforce these provisions. Dispute settlement provisions vary, but usually include arbitration procedures, often based in recent BITs on international standards, such those of the World Bank Group's International Center for Settlement of Investment Disputes (ICSID) or the UN Commission on International Trade Law. BITs involving the United States generally allow private investors to bring cases against host-country governments, thereby removing any foreign policy–driven uncertainty regarding the willingness of their home country governments to defend their rights.

BITs have proliferated sharply since the 1980s. The total number of BITs in force has risen from 355 in 1990 to 1,891 by the end of 2005; another 604 had been signed but not yet entered into force (UNCTAD 2006a). The pace of new annual signings decreased — from more than 200 in the mid-1990s to 70 in 2005 — as the pool of countries willing to enter into agreements began to reach full coverage. Although the majority of agreements are North-South arrangements, recent signings feature agreements among developing countries. South-South agreements now constitute 29 percent of the total. The top 10 signatories of BITs included seven from Europe (Germany, France, the United Kingdom, Italy, Belgium, the Netherlands, and Switzerland). China, Egypt, and Romania round out the top 10.

BITs now cover a substantial share of foreign direct investment (FDI) flows from members of the Organisation for Economic Co-operation and Development (OECD) to developing countries. In 1990, roughly 9 percent of investment flows were covered (though data in the early years are somewhat understated). By 2004, 43 percent of investment was covered.

Preferential Trade Agreements (PTAs) PTAs, like BITs, have multiplied in the years since 1990. The United Nations Conference on Trade and Development (UNCTAD) reports that 232 PTAs were in force at the end of 2005, 86 of which were purely South-South arrangements. Many of these agreements, particularly the North-South agreements, now cover investment. PTAs involving the United States and the European Union (EU) have been important drivers. By 2006, the United States had in force free trade agreements with investment provisions involving 14 countries, including Mexico, Chile, and Singapore, with three signed and before the U.S. Congress, and three more under negotiation. The EU engagement in PTAs is even more extensive, with more than 100 PTAs as of 2003, and investment is more frequently covered. Thus, for example, investment policies are being discussed as part of the Economic Partnership Agreement negotiations with African, Caribbean, and Pacific countries and the recent agreements with southern Mediterranean countries.

U.S. Free Trade Agreements (FTAs) Recent U.S. FTAs include preestablishment rights of market access. These provisions imply opening services markets to competition from foreign suppliers — or locking in prior autonomous liberalization — except in those sectors excluded (through a negative list). This therefore may expand on the coverage of com-

mitments in the General Agreement on Trade in Services (GATS), where a positive list is used and greatly enhances the transparency of prevailing policies. Since most of the countries with which the United States has concluded bilateral PTAs are already open in most sectors, the agreements lock in prevailing openness and generally effect changes in only a few still-restricted activities. Common provisions range from inclusion of insurance, financial advisory services, and selected telecommunications services to arguably relatively inconsequential changes to already open regimes, such as the commitment by Singapore to cease cross-subsidies in express mail delivery or the commitment of Chile in insurance services and a few other sectors. Moreover, notable for their absence is the exclusion of labor services, except provisional visas for professionals associated with investing firms.

All U.S. FTAs provide for national and MFN treatment. For many of the initial FTA countries, these had long been included in national legislation and/or have been incorporated into bilateral investment treaties on a postestablishment basis.

The U.S. FTAs have subsumed preexisting BITs and provided new measures covering investment. Agreements, especially those signed after the North American Free Trade Agreement (NAFTA), include broad definitions of investment, including not only FDI but also portfolio flows, private debt and even sovereign debt issues as well as intellectual property. Such broad definitions of investment policies can expose countries to dispute settlement across a wide range of assets.

The inclusion of intellectual property rights in the definition of assets covered by the investment provisions creates a potential liability for signatory governments. Not only are the intellectual property rights far more extensive under the recent PTAs than under the WTO's Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) (see Fink and Reichenmiller 2005), the dispute settlement provisions are more powerful. For example, if a government decides to issue a compulsory license to control drug prices and the patent owner disputes this action under the terms of the U.S. PTA, the patent holder can take the claim to commercial arbitration under the PTA's investment provisions. This instrument is considerably more powerful for intellectual property rights enforcement than the state-to-state provisions under the TRIPS Agreement (see World Bank 2003).

Another area of disciplines concerns trade-related investment measures (TRIMs) government policies that require foreign companies to export in a certain portion of their sales or balance trade, using local inputs to achieve value-added objectives. All of the bilateral U.S. PTAs ban TRIMs. The U.S. bilateral agreements have in effect established a "WTO TRIMs Plus" set of obligations that include outright bans on certain performance requirements, including management restrictions and export, minimum domestic content, domestic sourcing, trade balancing, and technology transfer requirements. Government procurement, environmental standards, and requirements for local research and development are generally not covered.

A key feature of U.S. agreements is the use of negative lists excluding sectors or industries from coverage. This feature implies that all other sectors not mentioned including new economic activities are included; this is distinct from the positive list approach typically found in EU agreements and most South-South agreements. Ratchet provisions, also a feature of U.S. agreements, bind liberalizing policies enacted subsequent to the agreement's entrance into force, thus locking the policy framework progressively in ever greater openness and certainty.

Finally, with the exception of the Australian FTA, all the U.S. agreements contain an investor-state dispute resolution provision that permits investors to take foreign governments to dispute resolution for violation of the treaty's national treatment, nondiscrimination, or expropriation provisions, among others.

European Union Agreements Because investment is the competence of the EU member states, not the European Commission, EU PTAs have treated investment only generally or indirectly. The earliest (and least comprehensive) are the Euro-Mediterranean Partnership agreements (starting in 1995) and

the South African agreement (1999), which contained virtually no investment provisions. Thus the market access provisions were general, directed mainly at services, and contain only the promise of potential liberalization after discussions to transpire some 5 years after the entry into force of the agreement. In the EU-Mexico PTA, several general provisions were included, many ratifying GATS arrangements, as well as specific liberalization commitments in the financial sector. The EU-Chile agreement goes further by additionally locking in some liberalization of telecommunications and maritime services.

The EU agreements with Mexico and Chile, though more comprehensive than earlier agreements, do not have the same strength of disciplines as the U.S. agreements. The trade provisions use a positive list and implicitly exclude new products. The treatment of investment and capital flows in both agreements were not extensive. For example, the EU-Mexico agreement simply states that the existing restrictions on investment will be progressively eliminated and no new restrictions adopted, without specifying particular sectors or setting a timeline for liberalization. The language in the EU-Chile agreement is even more general, calling for "free movement of capital relating to direct investments made in accordance with the laws of the host country." Both agreements allow for use of safeguards in the event of monetary or exchange rate difficulties, and although the time limit is set at 6 months for Mexico and 12 months for Chile, allow for continuation of the safeguard after the time limit through its formal reintroduction.

The treatment of dispute settlement involves state-to-state mechanisms rather than the investor-state provisions in U.S. agreements that allow private companies to initiate arbitration cases against governments. The EU provisions pertaining to investment are subsumed in the general dispute settlement provisions for all matters in the PTA. State-to-state dispute settlement is first attempted through consultations with a Joint Committee (Association Committee in the case of Chile and Mediterranean countries) within 30 days of a party's request. If this step of "dispute avoidance" proves unsuccessful, in the case of Chile, the concerned party can forward its request to an arbitration panel comprising representatives of both parties. The arbitration panel's decisions are binding, and the panel can also rule on the conformity of any measures undertaken as a result of its decision with the original ruling. Both agreements provide extensive detail on the process of appointing members to the arbitration panel, timelines for the panel's ruling, and compliance with the panel's decisions. Still, companies in member states that have their own BITs with recipient countries can activate investor state arbitration through BIT mechanisms.

Multilateral Initiatives Since 1948 when the Havana Charter was created, governments in high-income countries have sought to weave investment provisions into international trade law. While selected investment issues were addressed with the creation of the UN Center on Transnationals and the OECD Guidelines for Multinationals, it was not until the OECD launched the Multilateral Agreement on Investment in 1995 that a full-scale effort was resumed. This effort eventually foundered in the face of resistance from labor and environmental groups as well as developing countries that were largely excluded; when France withdrew from the negotiations in 1998 over the failure to protect its media and cultural interests, the negotiations collapsed.

With the launching of the Doha Development Agenda in the Ministerial Meeting of the WTO in November 2001, it seemed that investment would once again be taken up as part of the negotiating mandate for the so-called Singapore issues. Discussions in the 20 subsequent months before the next Ministerial meeting in Cancun, Mexico, in 2003, focused on the definition of investment to be covered, whether to adopt positive or negative lists, and the type of dispute resolution to be adopted. In each of these areas, the contentious negotiations tended to gravitate toward outcomes favoring the lowest common denominator: narrowly focused definitions of investment, positive lists with abundant room for exclusions, and state-to-state dispute resolution.

However, progress in the Doha Round was painfully slow in the run-up to the Cancun meeting, and disagreement on the Singapore issues was only one symptom. The inclusion of investment in the negotiations encountered vehement opposition from a coalition of developing countries, including most notably India and Malaysia. Negotiations were stymied over other major issues in agriculture and nonagricultural products, and many developing countries considered the heated discussions on the Singapore issues a distraction and of little value to their development.

Finally, in an ill-fated effort to move the discussions off dead center, the EU, the principal *demandeur* of Singapore issues, at the last minute offered to abandon investment and competition and even cease future WTO activity in the Working Group. The Cancun Ministerial nonetheless broke down in acrimony. Only later when India offered to accept one of the four Singapore issues, trade facilitation, did this later become the accepted position in subsequent negotiations. Investment was officially dead as an issue for negotiation in the WTO's Doha Round.

Effects on Investment A main objective of a BIT is to increase the flow of investment to signatory countries. Does the signing of bilateral investment treaties in fact increase the flow of FDI? Hallward-Driemeier (2003), in one of the first systematic statistical analyses, considered bilateral flows from OECD members to 31 developing countries over two decades. Her analysis found that, controlling for a time trend and country-specific effects, BITs had virtually no independent effect in increasing the share of FDI to a signatory country from a home country. Countries signing a BIT were no more likely to receive additional FDI than countries without such a pact. Even comparing flows in the three years after a BIT was signed to the three years prior, there was no significant increase in FDI. Rose-Ackerman and Tobin (2005) analyzed the flows to 63 countries with data averaged over a five-year period and found that only with countries at very low levels of risk was there any increase of investment from the signatory home country; in a detailed analysis of U.S. FDI behavior,

they failed to find any statistically significant effects on U.S. FDI to 54 countries, irrespective of political risk.

In contrast to these findings, Neumayer and Spess (2005), using a larger sample over a longer period and considering simply the number of BITs signed, found robust and positive effects of the total number of BITs on total inflows of FDI. This finding relies heavily on the notion that signing a BIT signals to all would-be investors that the environment is more welcoming. In this model, the signal is more important than the specific investor protections. These contradictory findings at minimum suggest that the jury is still out and that it cannot be accepted axiomatically that BITs increase investment flows.

Assessing the evidence of investment provisions in PTAs is more difficult. The reason is that it is difficult to distinguish the effects on investment flows of creating a larger, single market from the effects of investor protections and other investment provisions in the trade agreement. Empirical studies generally consider the combined effects of PTAs trade and investment provisions together.

Available evidence suggests mild positive effects on overall investment. The Global Economic Prospects 2005 (World Bank 2004) examined the effects of PTA membership and other variables on FDI inflows for a panel of 152 countries during the 1980 2002 period. The study found that PTAs that result in larger markets do attract greater FDI. The interaction of signing a PTA and expanded market size associated with the integrated markets is significant and positively related to FDI. On average, a 10 percent increase in market size associated with a PTA produces an increase in FDI of 5 percent.

One example is NAFTA, which is a comprehensive arrangement that includes significant investor protections in combination with broad-based tariff reductions and border liberalizations. Chapter 11 of the NAFTA agreement allows investors to sue the government in the event of regulatory or other actions that might diminish the value of a foreign investment. In their study, Lederman, Maloney, and Serven (2003) find that, in addition to positive forces

in the global economy that propelled investment into Mexico and other emerging markets after 1994, the trade opening and NAFTA accession also played a role in Mexico's FDI rise.

Neither this study nor others attempt to distinguish the role of enhanced investor protections from access to the Mexican market and its other resources in increasing the flow of FDI. It may be that simply liberalizing the investment law in 1993 and making it easier to take advantage of productivity-adjusted wage differentials in the NAFTA countries was sufficient to explain the difference. However, it is possible that investor protections in chapter 11 played an important independent role because Mexico's legal framework was seen as less reliable than other investment destinations. In an analysis of shareholder rights, creditor rights, efficiency of judiciary, rule of law, and absence of corruption, Mexico scores below the Latin American average in four of five measures (as cited in Lederman, Maloney, and Serven 2003). To the extent that chapter 11 provided investors with additional comfort over and above the existing investment climate, its protections would have offset these disadvantages. In any case, Lederman et al. (2003) conclude that Mexico's entry into NAFTA led to an increase in annual FDI by around 40 percent.

Dispute Settlement: Rising Case Load and Rising Awards The costs of resolving investor disputes can be large. The number of cases brought under the various bilateral investment treaties and PTAs has risen dramatically. Though the absence of reporting data from all the dispute tribunals prevents full analysis, the International Center for Settlement of Investment Disputes (ICSID) at the World Bank reports a steady increase in the number of cases from an annual average of 1.5 cases in 1972 95 to 29 in 2003 4, after which a tapering off is evident. However, UNCTAD has established a database that includes other sources, and its main report (UNCTAD 2005) and subsequent update (UNCTAD 2006b) point to a continued upward trend through 2005. Some 48 known cases were brought in 2005 (UNCTAD 2006b).

In fact, the dispute settlement process is cloaked in opaqueness. The ICSID registry contains only minimal information about the dispute it is not known whether the dispute arises from a BIT provision, a FTA provision, or some other state contract. Information often becomes available only when one of the parties to an arbitration makes it public. Many of the disputes outside of the ICSID tribunal are never made public, and even less is known about their resolution.

United States FTAs do provide for somewhat greater transparency. In NAFTA, for example, as of July 2004, there were 31 cases brought under chapter 11 (including 14 against Mexico, 9 against Canada, and 8 against the United States). Six cases have been decided in favor of the investor, but the amount awarded has been small compared to initial (undoubtedly inflated) claims. Tribunal awards have totaled $35 million compared to claims of $1.0 billion.

At least 61 governments 47 from the developing world (including transition countries) have faced investment treaty arbitration. Forty-two cases were lodged against the Argentine government after its devaluation of the peso in 2001. Causes have varied from changes in tax policy perceived as adverse by investors to expropriations following conflict or coups, irregularities in bidding processes, and others (Peterson 2003a, 2003b).

Awards can be hefty. In February 2005, the Slovak Republic agreed to comply with an ICSID arbitral award of $830 million to CSOB, a Czech bank. This award surpassed that of a tribunal in Stockholm, which required the government of the Czech Republic to pay one company, Central European Media (CME), $350 million for violation of a BIT that deprived CME of a stake in an English-language TV station in Prague. This amount was ten times higher than previously known awards under arbitration cases and about equal to the entire public sector deficit of the Czech Republic (Peterson 2003a, 2003b). But the biggest case has been filed against the Russian Federation, involving a claim for $28 billion for the alleged expropriation of Group Menatep's

majority shareholding in the Yukos oil firm under the 1994 Energy Charter Treaty (as yet unratified). In one of a series of decisions that have gone against the government of Argentina, in February 2007 Siemens was awarded $217 million for violation of a Germany-Argentina bilateral investment treaty (Peterson 2007). To these awards must be added legal fees and tribunal costs. Governments may have to pay costs that can range from $1 million to $2 million (UNCTAD 2005).

Flaws in the System of Property Rights Enforcement Theodore Moran (2006) discusses weaknesses in the dispute settlement procedures through the lens of prominent NAFTA cases. Perhaps the most detailed case study of investment dispute resolution and its consequences is the Wells and Ahmed (2007) study of Indonesia in the wake of the 1998 East Asia crisis. In an illuminating history of the collapse of private power contractual arrangements and subsequent dispute resolution, Wells and Ahmed conclude that the arbitration process "served the host countries poorly" and that companies found the promised protections were "largely illusory." Wells and Ahmed point to five weaknesses:

1. The system of contracts and enforcement is overly rigid, failing to take into account changed circumstances, excessively large claims, the role of corruption, and reasonable expectations from public policy.

2. The system produces inconsistent outcomes, as "tribunal shopping" has led to wildly different outcomes based on differing standards of evaluation using the same contracts.

3. The investor-state arbitration system is inherently asymmetric, according investors a superior position in the resolution process, in relation to both governments (that cannot initiate a case against an underperforming investor) and to domestic investors (who cannot avail themselves of the system).

4. The system creates a "moral hazard" that effectively discourages companies, based on rigidly interpreted rights, from seeking res-

olution through renegotiating contractual arrangements.

5. The system, by focusing only on awarding damages, foregoes the opportunity to encourage mutually beneficial contractual restructuring, policy changes, or negotiated settlements.

Many of these conclusions also emerge from the study of Cosbey, Mann, Peterson, and Von Molke (2004) who examine investment agreements from a cross-country perspective. They underscore the lack of transparency, potential conflicts of interests in the selection of arbitrators, inconsistency across multiple tribunals reviewing the same case, and lack of accountability of the system. Moran (2006) detects a lessening of salience of treaty-driven investor protections as experience tends to caution investors against overrelying on these instruments and moderate expectations of developing countries about higher investment flows.

In conclusion, international investment agreements are on the rise, even though their ability to promote investment remains unclear. Several questions merit additional exploration: What recourse do investors have legally if a government refuses to pay claims awarded against it, and does recourse provide sufficient incentive for full payment? Because the costs of enforcing payment of an award are usually high, governments could negotiate down the final settlement. What has been the experience of investors after being awarded large claims; have they collected full amounts, and if not, what was the negotiating process that led to final settlement?

Further, how can the arbitration process be made more transparent, fairer, and more open? Is there a synergy between investor protections and free trade provisions that would encourage FDI? Do countries with negative list provisions in their BITs and FTAs attract more investment than countries with positive list provisions?

If the objective is to increase the inflows of FDI, countries should begin with improvements in the domestic investment climate sound macroeconomic policies, stable property rights for all investors,

and sound regulatory policies and treat investment treaty provisions as a complement rather than a substitute for an attractive policy framework.

See also domestic content requirements; foreign direct investment and international technology transfer; foreign direct investment and tax revenues; foreign equity restrictions; free trade area; trade costs and foreign direct investment; trade-related investment measures (TRIMs)

FURTHER READING

Cosbey, Aaron, Howard Mann, Luke Peterson, and Konrad Von Moltke. 2004. *Investment and Sustainable Devel opment: A Guide to the Uses and Potential of International Investment Agreements.* Ottawa: International Institute for Sustainable Development. A comprehensive guide to international agreement provisions.

Fink, Carsten, and Patrick Reichenmiller. 2005. "Tight ening TRIPS: Intellectual Property Provisions of U.S. Free Trade Agreements." In *Trade, Doha, and Devel opment: A Window into the Issues*, edited by Richard Newfarmer. Washington, DC: World Bank, 187 97. One of the first detailed analyses of IPR provisions in U.S. FTAs and their impact on access to medicine.

Hallward Driemeier, Mary. 2003. "Do Bilateral Invest ment Treaties Attract FDI? Only a Bit . . . and They Could Bite." World Bank Policy Research Paper No. 3121. Washington, DC: World Bank. A sophisticated econometric study of the effects of BITs on investment flows.

Lederman, Daniel, William F. Maloney, and Luis Serven. 2003. *Lessons from NAFTA for Latin America and Ca ribbean*. Washington, DC: World Bank. The most de tailed study of NAFTA to date.

Moran, Theodore. 2006. "Toward a Development Friendly International Regulatory Framework for For eign Direct Investment." (September). Washington, DC: mimeo. A thoughtful discussion of ways to improve development outcomes through multilateral regulation.

Neumayer, Eric, and Laura Spess. 2005 "Do Bilateral In vestment Treaties Increase Foreign Direct Investment to Developing Countries?" *World Development* 33 (10): 1567 85. Another serious study of BITs and investment flows to developing countries.

Peterson, Luke. 2003a. "Research Note: Emerging Bilateral Investment Treaty Arbitration and Sustainable Devel

opment." IISD Invest New Bulletin (April). Winnipeg, Canada: International Institute for Sustainable Devel opment. Renamed the Investment Treaty News, this electronic newsletter is an invaluable source of current information on investment disputes and related devel opments.

. 2003b. "Czech Republic Hit with Massive Compensation Bill in Investment Treaty Dispute." IISD Invest New Bulletin (March 21). Winnipeg, Canada: International Institute for Sustainable De velopment.

. 2007. "Argentina Liable for \$217 Million in In vestment Treaty Arbitration with Siemens." *Investment Treaty News* (February 19). Available at http://www.iisd .org. Winnipeg, Canada: International Institute for Sus tainable Development.

Rose Ackerman, Susan, and Jennifer Tobin. 2005. "For eign Direct Investment and the Business Environment in Developing Countries: The Impact of Bilateral Invest ment Treaties." Yale Law and Economics Research Pa per 293 (May). New Haven: Yale University. A useful analysis of BITs and investment.

United Nations Conference on Trade and Development (UNCTAD). 1998. *Bilateral Investment Treaties in the Mid 1990s*. New York: United Nations. UNCTAD is the leading data source for information on FDI around the world.

. 2005. "Investor State Disputes Arising from In vestment Treats: A Review." Geneva: UNCTAD.

. 2006a. "Entry into Force of Bilateral Invest ment Treaties (BITs)." IIA Monitor No. 3. Geneva: UNCTAD.

. 2006b. "Latest Developments in Investor State Dispute Settlement." IIA Monitor No. 4. Geneva: UNCTAD.

Wells, Louis, and Ahmed, Rafiq. 2007. *Making Foreign Investment Safe: Property Rights and National Sovereignty.* New York: Oxford University Press. One of the most in depth political economic studies of investment disputes and their causes, resolutions, and weaknesses.

World Bank. 2003. *Global Economic Prospects 2004: In vesting to Unlock Global Opportunities*. Washington, DC: World Bank. Examines investment policy from a development perspective, with a chapter on investment in WTO negotiations.

. 2004. *Global Economic Prospects 2005: Trade, Regionalism, and Development.* Washington, DC: World Bank. Provides a comparative analysis of investment provisions in different settings from U.S. and EU agreements.

RICHARD NEWFARMER

International Labor Organization

The International Labor Organization (ILO) was created in the wake of World War I and the communist takeover of Russia with the narrow goal of improving working conditions around the world. The underlying objectives, however, were to avoid political instability by addressing issues of social justice and concerns that global competition could undermine countries' efforts to improve working conditions at home. The first wave of globalization had already crashed on the shoals of World War I by the time the ILO had come into being. But a century later, a new wave of globalization triggered similar concerns and the ILO once again found itself at the center of debates over how to address economic insecurity and inequality.

The ILO is the only League of Nations institution to survive World War II and is also the only international organization that is not purely intergovernmental in its governance structure. Union and employer group representatives are part of each country's delegation and have the same right to vote as government representatives at policymaking meetings. As of 2006, the ILO had adopted 187 conventions determining international standards for various aspects of work and employment. Some of the standards dealt with fundamental rights of workers, but most are narrow and technical, on issues of importance only to labor, management, or regulators in specific sectors, for instance, shipping or communications.

In 1998, the ILO approved a "Declaration on Fundamental Principles and Rights at Work," which provided a consensus definition of four core labor standards that became the centerpiece of the global standards movement. In the face of accelerating globalization, the ILO has also focused more attention on employment, adjustment, and migration issues. In 2002, it established a World Commission on the Social Dimensions of Globalization to explore ways to make economic globalization more inclusive and to spread its benefits more broadly.

The Organization of the ILO The ILO is a tripartite organization with 179 member states and 716 voting delegates, with each member having two government representatives and one each representing employers and workers. In theory, worker and employer delegates are not bound by their government's position and vote independently. Delegates gather annually at the International Labor Conference, where new conventions may be adopted, implementation of existing conventions is reviewed, and the budget is approved. The Governing Body, except for ten permanent members, is elected by the conference and is the executive body of the organization, responsible for developing policy, electing the director-general, and overseeing the organization's work program.

The International Labor Office, headed by the director-general, is the secretariat of the organization with headquarters in Geneva and branch offices around the world. The office provides advisory services and technical assistance, for example on how to create or reform laws for social protection, and carries out research projects, such as studying the implications of the end of the Multifiber Arrangement for workers in clothing factories around the world. The activities of the office are organized around the theme of "decent work" and four strategic objectives: rights at work, employment, social protection, and social dialogue.

The ILO's Tools In addition to standards-setting, the ILO has three main tools for improving working conditions. It supervises compliance with global labor conventions and publicizes violations of standards to shame countries into improving. It gives technical assistance to labor ministries and other agencies, unions, and employer groups to address employment-related issues and to improve the implementation of labor standards. And it can punish

countries that do not comply with their commitments, though complaints have rarely led to formal enforcement measures.

Supervising and Publicizing Country Performance
The ILO has extensive mechanisms for supervising the application of conventions, including routine reporting by countries on implementation of conventions they have ratified, why they have not ratified others, and what they are doing to achieve the goals of those conventions. These reports are reviewed by independent experts who identify areas both of progress and of special concern, which may be taken up by the Conference Committee on the Application of Standards at the annual conference.

The most important innovation in ILO monitoring came with the 1998 Declaration on Fundamental Rights and Principles at Work, which defined four core labor standards:

- freedom of association and the right to organize and bargain collectively;
- freedom from forced labor;
- the eventual abolition of child labor;
- nondiscrimination in employment.

The follow-up mechanism for implementing the declaration requires member countries that have not ratified one or more of the eight conventions associated with these principles to report annually on what they are doing to promote the conventions and encourages employer and worker groups to comment on the national submissions. In addition, the conference requires the director-general to prepare a global report summarizing how each core standard is being implemented around the world and identifying key obstacles to compliance.

Technical Assistance Ministries of Labor and other organizations seeking to raise standards and improve the operation of labor markets are often weak bureaucracies with inadequate resources and limited support from political leaders. In such situations, even modest ILO technical assistance can be helpful. The ILO also conducts extensive research and outreach on issues related to employment and underemployment in developing countries, as well as the obstacles to standards compliance posed by extensive employment in the informal sector in many countries.

For decades ILO technical assistance programs received only modest funding from wealthier member states and the agency had a limited presence in developing countries. In the 1990s, however, many countries, including the United States, responded to demands that the world community do more to improve labor standards by increasing their funding for ILO programs. As a result, technical cooperation spending increased from an average of just over $90 million per year in 1998 2000 to nearly $140 million in 2004. Virtually all of the increase went to programs to eliminate child labor, which accounted for nearly 40 percent of total technical assistance in 2004.

Enforcement of Conventions and Standards In addition to routine reporting and review, the ILO has two ad hoc mechanisms for promoting compliance with labor standards, though critics charge that the organization is toothless. Article 24 of the ILO Constitution allows any worker or employer organization in the world to file a complaint alleging that a member government is not complying with a convention it has ratified or that a country has violated freedom of association (regardless of whether it ratified the relevant conventions). If such complaints are not resolved through informal consultation, the Governing Body can refer them to the Committee on Freedom of Association or appoint an ad hoc committee to analyze the situation, ask the government to respond, and make recommendations on what the government can do to comply. If the problem remains unresolved, official ILO delegates can file an article 26 complaint.

The ILO Governing Body will again try to resolve the problem informally through consultations. But if that leads nowhere, it can appoint a Commission of Inquiry to formally investigate the charges and recommend how to correct the problem. The target of the complaint can appeal the commission's finding to the International Court of Justice. If the commission's findings are upheld and a satisfactory resolution is still not forthcoming, article 33 of the ILO

Constitution provides that "the Governing Body may recommend that the Conference take such action as it may deem wise and expedient to secure compliance therewith."

Between 1919 and 1960, there was only one article 26 complaint. From 1961 to 2000, there was an average of six complaints per decade. In its history, the ILO has appointed only nine Commissions of Inquiry, usually regarding alleged violations of one of the four fundamental rights. The ILO did not take the next step of invoking article 33 until 2000, when the conference approved a resolution calling on member states "to review their relationship with the Government of Myanmar (Burma) and to take appropriate measures to ensure that Myanmar 'cannot take advantage of such relations to perpetuate or extend the system of forced or compulsory labour.'" Raising questions about the ILO's effectiveness in difficult situations, however, the annual conference asked the Governing Body to consider what additional steps might be taken if the government in Myanmar remained uncooperative.

In sum, the ILO aims to improve living conditions for people around the world by increasing employment opportunities, making work safer and more productive, and in recent years, ensuring that workers share in the benefits of globalization. It does this by promoting social dialogue and basic rights that allow workers to bargain over wages and working conditions; by promulgating, monitoring, and enforcing other standards; and by providing technical assistance to employers associations, unions, and governments in pursuit of these objectives. The ILO also conducts research on a variety of employment-related issues and is the major source for what data exists on labor markets.

See also child labor; labor standards; trade and wages

FURTHER READING

Charnovitz, Steven. 2000. "The International Labor Organization in Its Second Century." In *Max Planck Yearbook of United Nations Law,* edited by Jochen A. Fruwein and Rudige Wolfram. New York: Kluwer, 147 84. Analyzes the challenges and opportunities facing the ILO in a world with a more globalized economy and greater involvement by developing countries.

International Labor Organization. http://www.ilo.org. The ILO's Web site provides an immense amount of in formation about the organization's history and current operations.

Sengenberger, Werner, and Duncan Campbell. 1994. *In ternational Labor Standards and Economic Inter dependence: Essays in Commemoration of the 7th Anni versary of the International Labor Organization and the 50th Anniversary of the Declaration of Philadelphia.* Geneva: International Institute for Labor Studies. Cov ers the history and economics of international labor standards, as well as issues such as new challenges from the latest wave of globalization and whether or which standards should be universal.

World Commission on the Social Dimensions of Globali zation. 2004. *A Fair Globalization: Creating Opportu nities for All.* Geneva: International Labor Organization. Report of an independent commission chaired by the presidents of Finland and Tanzania and including rep resentatives from governments, nongovernmental or ganizations, business, and labor.

KIMBERLY ANN ELLIOTT

■ international liquidity

International liquidity refers to the availability of internationally accepted means of settling international debts relative to the demand or potential demand for such financial assets. The availability or supply of international liquidity depends on the stock of total reserve assets owned by world central banks or national monetary authorities, as well as the ability of these institutions to borrow from the International Monetary Fund (IMF), the Eurocurrency market, and other financial institutions. The adequacy of international liquidity also depends on the demand for, or the expectation of, having to make such payments.

International reserve assets comprise reserve currencies, Special Drawing Rights (SDRs), and gold owned by nations' central banks or monetary authorities, as well member nations' reserve

position in the IMF. Reserve currencies are foreign currencies held by central banks or monetary authorities that can be used (i.e., are accepted) for the settlement of intergovernmental claims. Reserve currencies represent almost 90 percent of total international reserve assets and the U.S. dollar accounts for about 65 percent of the total stock of reserve currencies. The second most important reserve currency is the euro. Other reserve currencies are the Japanese yen, the British pound, and the Swiss franc.

Special Drawing Rights (SDRs) are accounting entries on the books of the IMF introduced in 1967 to supplement international reserves. SDRs are used only among central banks to settle balance-of-payments deficits or surpluses; they cannot be used in private commercial dealings. At the 1967 IMF meeting it was agreed to create SDRs in the amount of $9.5 billion to be distributed to member nations according to their quotas in the IMF in three installments in January 1970, 1971, and 1972. Further allocations of SDRs made in the 1979 81 period brought their total to about $19 billion. The value of one SDR was originally set equal to one U.S. dollar, but rose above $1 as a result of the devaluations of the dollar in 1971 and 1973.

Starting in 1974, the value of SDRs was tied to a basket of currencies.

Gold was the international reserve asset par excellence during the gold standard period (1870 1914). Under the Bretton Woods system of fixed exchange rates (1947 71), gold was one of several types of international reserve asset held by national central banks or monetary authorities and the United States used gold as backing for the dollar reserves held by foreigners. In 1971, the United States suspended the dollar convertibility into gold at the fixed price of $35 per ounce. The price of gold then increased on the world market, but monetary gold (the amount held by nations' central banks or monetary authorities) continued to be valued at $35 per ounce in official transactions. Today, gold has lost a great deal of its importance as an international reserve asset and the IMF is moving toward demonetizing it, so that in the future gold would be used only for commercial purposes.

Member countries' reserve position in the IMF also constitutes a part of international reserves. This is equal to the amount of international reserves (originally, gold and reserve currencies but now only reserve currencies) that nations pay into the fund (based on their quota into the fund) on becoming

Table 1
International reserves, 1950 2005 (billions of SDRs, at year end)

International reserve asset	1950	1960	1970	1980	1990	2000	2005
Foreign exchange	13.3	18.5	45.1	292.6	611.3	1,490.2	2,918.2
SDRs			3.1	11.8	20.4	18.5	20.1
Reserve position in the fund	1.7	3.6	7.7	16.8	23.7	47.4	28.6
Total reserves minus gold	15.0	22.1	56.2	321.3	655.4	1,556.1	2,966.8
Gold at SDR 35/ounce	32.2	37.9	37.0	33.5	32.9	33.3	30.9
Total with gold at SDR 35/ounce	48.2	60.0	93.2	354.7	688.3	1,589.4	2,997.7
Gold at SDR market price	33.0	38.6	39.6	455.4	253.1	200.6	316.6
Total with gold at market price in SDR	48.0	60.7	95.8	776.6	908.3	1,756.7	3,283.4
U.S. dollar per SDR	1.0000	1.000	1.000	1.2754	1.4227	1.3029	1.4293

Source: IMF, *International Financial Statistics Yearbook,* various issues.

members, and which nations can automatically borrow from the fund without any restrictions or questions asked.

The amount of the various international reserve assets available to nations from 1950 to 2005 (at ten-year intervals) is given in table 1, with and without the inclusion of gold reserves, and with gold at the official price of SDR 35/ounce and at market price. Note that the amounts of the various international reserve assets are given in SDRs by the IMF, but their dollar value can be obtained by multiplying them by the dollar value per SDR shown in the last row of the table. For example, the foreign exchange reserves of SDR 2.9185 trillion at the end of 2005 was equal to $4.171 trillion (SDR 2,918.2 billion times the dollar value of the SDR of $1.4293/SDR in 2005).

The adequacy of international liquidity depends not only on the supply or availability of liquidity but also on the demand for liquidity. Under the Bretton Woods system, nations needed international reserves to pay for temporary balance-of-payments deficits, without the need to restrict imports directly or indirectly through slower domestic growth. During this period, most of the increase in world liquidity resulted from the increase in official holdings of foreign exchange reserves, mostly dollars, to finance U.S. balance-of-payments deficits. The problem was that the more U.S. deficits persisted and the more dollars accumulated in foreign hands, the less willing foreigners were to hold additional dollars and the less confidence they had in the dollar (i.e., that the dollar would retain its international value and that the United States would be able to honor its commitment to redeem dollars in gold at the fixed price of gold of $35 per ounce, in the face of dwindling U.S. gold reserves). Indeed, it was precisely to avoid this confidence problem that the IMF decided to create SDRs in 1967. But this was not enough to prevent the collapse of the Bretton Woods system in 1971, when some central banks demanded conversion of some of their dollar reserves into gold, thus triggering the collapse of the system.

After the collapse of the Bretton Woods system, most developed nations and many developing nations moved to a system of managed exchange rate flexibility, which persists to this day. Under such a system, a nation with an actual or potential balance of payments deficit would experience a depreciation of its currency, which would correct the deficit by stimulating the nation's exports and discouraging its imports. Since the adjustment takes time, however, nations continue to need international reserves or liquidity. Nations also use international reserves or liquidity to intervene in foreign exchange markets to smooth out excessive exchange rate volatility or when attempting to influence the underlying value of the exchange rate of their currencies. Central banks could also increase their interest rate to attract international capital flows to cover balance of payments deficits or borrow additional reserves from the IMF or in the Eurocurrency market at a charge. In the case of currency crises (such as those that afflicted several emerging markets economies from 1994 to 2002), liquidity could be mobilized on an international scale under IMF leadership.

See also balance of payments; Bretton Woods system; currency crisis; dominant currency; Eurocurrencies; foreign exchange intervention; gold standard, international; international financial architecture; International Monetary Fund (IMF); international reserves; money supply; reserve currency; special drawing rights; vehicle currency

FURTHER READING

Bank of International Settlements. *Annual Report*. Basel: BIS. A report on the world's financial and banking situation.

De Vries, M. G. 1986. *The IMF in a Changing World*. Washington, DC: IMF. An authoritative discussion of the operation of the International Monetary Fund from its creation after World War II until the early 1980s.

International Monetary Fund. Annual. *International Financial Statistics Yearbook*. Washington, DC: IMF. The most complete and comparable financial data of member nations of the International Monetary Fund.

Mundell, Robert A. 2000. "A Reconsideration of the Twentieth Century." *American Economic Review* (June): 327 40. A review and interpretation of the operation of the international monetary system during the past century by a Nobel Prize winner in economics.

Salvatore, Dominick. 2007. *International Economics*. 9th ed. Hoboken, NJ: Wiley, chapter 21. An examination and

evaluation of the operation of the international mone
tary system from the time of the gold standard to the
present and a discussion of its possible reforms in the
future.

Triffin, Robert. 1961. *Gold and the Dollar Crisis*. New
  Haven, CT: Yale University Press. A classic of the early
  postwar trouble of the dollar resulting from the "dollar
  overhang," or the excessive supply of dollars in the hands
  of foreign monetary authorities.

Willett, Thomas D. 1980. *International Liquidity Issues*.
  Washington, DC: American Enterprise Institute. A
  discussion of the adequacy of international liquidity re
  quired for the smooth operation of the international
  monetary system.

DOMINICK SALVATORE

■ International Monetary Fund (IMF)

The International Monetary Fund (IMF) is an
intergovernmental organization that seeks to ad-
vance global financial stability. The IMF's respon-
sibilities, as set out in its Articles of Agreement,
include promoting international monetary coop-
eration and exchange stability, facilitating the ex-
pansion and growth of international trade, and
making its resources available to members experi-
encing balance of payments difficulties. There are
185 member nations, and the headquarters is in
Washington, DC.

The IMF was established in 1944 at the Bretton
Woods Conference in New Hampshire. The repre-
sentatives of 44 nations also created the International
Bank for Reconstruction and Development (World
Bank) for the purpose of postwar reconstruction; its
mandate was later expanded to encompass economic
development. The chief architects of the new inter-
national monetary system were Harry Dexter White
of the U.S. Department of the Treasury and the
noted British economist John Maynard Keynes.
They sought to design a system that would prevent
repetition of the chaos of the interwar period, when
countries abandoned the gold standard and depre-
ciated the values of their currencies in attempts to use
exports to bolster their economies.

In the Bretton Woods system, central banks
maintained fixed rates for their currencies in relation
to the U.S. dollar. They held their foreign currency
reserves in the form of dollars, which they could ex-
change for gold from the United States. The IMF
served as a monitor of the system to ensure that
governments fulfilled their obligations. The IMF also
provided credit to members with balance of pay-
ments deficits in order to provide time for their
governments to implement policies to restore equi-
librium.

The Bretton Woods system collapsed in the early
1970s, in part because foreign central banks feared
the inflationary consequences of the flow of dollars
from the United States. However, the IMF acquired
new roles by dealing with the debt crisis of the 1980s
and then the emergence of the transition economies
at the end of that decade. In the 1990s it coped with
international financial crises, such as those that oc-
curred in 1994 in Mexico and in East Asia in 1997,
while promoting growth in its poorest members.

Governance The IMF is under the jurisdiction of
its Board of Governors, which includes one member
and one alternative member appointed by each
country, and meets once a year. The Board of Gov-
ernors delegates the responsibility for supervising the
daily operations of the IMF to the Board of Direc-
tors. A country's position and voting power on the
Board of Directors is based on the size of its quota. A
country's quota is determined by its economic size
and the value of its international transactions. The
United States currently has the largest quota share,
17.09 percent, while Palau has the smallest relative
quota, .001 percent. Since major decisions require a
special majority of 85 percent of the votes cast, the
United States has an effective veto.

There are 24 executive directors representing the
IMF's members. The five countries with the largest
quotas appoint executive directors to represent
their interests on the board, while smaller members
form constituencies and elect a joint director. The
members with the largest quotas and their own
directors are the United States, Japan, Germany,
France, and the United Kingdom; Saudi Arabia, a
major creditor to the fund, China, and Russia form

single member constituencies and also name their own directors.

The Board of Governors conducts quota reviews at regular intervals, usually every five years. The Twelfth General Review took place in 2003, without any change in the quotas, and the Thirteenth General Review was scheduled to be completed by 2008. In the period between the 2003 and 2008 reviews, many members believed that the relative allocation of quotas was skewed in favor of some European nations at the expense of Asian economies and contended that there should be a reallocation of quota shares.

A country's quota also determines the amount of money it contributes to the IMF, called its subscription, and the limits on its access to financing. When it joins the IMF, a member makes a payment equal to its quota; one quarter of this subscription must consist of widely accepted currencies, such as the U.S. dollar or the euro. Increases in quotas require increases in the subscriptions paid in by the member nations. The quota payments provide the IMF the financial resources for its lending operations. In 2006 the IMF had $175 billion available for loans over the next one-year period.

The executive directors appoint a managing director to administer the fund's activities. The managing directors have always been Europeans, and Dominique Strauss-Kahn of France became the IMF's tenth managing director in 2007. There are three deputy directors to assist the managing director, and approximately 2,700 staff members, of whom about two-thirds are economists. The IMF has five area departments (African, Asia and Pacific, European, Middle East and Central Asia, and Western Hemisphere), eight functional departments (Finance, Fiscal Affairs, IMF Institute, Legal, Monetary and Capital Markets, Policy Development and Review, Research, and Statistics), and several information and support services departments. There is also an Independent Evaluation Office, which reports to the Executive Board.

Special Drawing Rights Special Drawing Rights (SDRs) are a unit of account used by the IMF. In the 1960s, central bankers outside the United States became concerned about the use of U.S. dollars as international reserves. The IMF's Board of Governors amended the original Articles of Agreement in 1969 to allow the introduction of the SDR by the IMF as a supplementary reserve asset. The first general distribution of SDRs was made in 1970 72. However, the need for new reserves diminished with the collapse of the Bretton Woods system and the development of private capital markets. The IMF continues to use SDRs as a monetary measurement. The value of the SDR is based on the foreign exchange value of the dollar, the euro, the yen and the British pound, and is posted on the fund's Web site (www.imf.org) daily.

Surveillance The IMF engages in surveillance activities to monitor and guide the economic policies of its members. The fund conducts bilateral surveillance of its individual members, as authorized under Article IV of the Articles of Agreement. A group of its economists, known as a staff mission, travel to a member's capital to consult with government officials about the member's exchange rate and monetary and fiscal policies. In recent years the talks have broadened to include the financial sector and assessments of vulnerability to capital flows, as well as structural issues such as fiscal reforms.

The IMF's economists submit a report of their findings to the fund's Executive Board, which uses it as the focus of its discussion of the country. Approximately nine out of ten of the IMF's members allow summaries of the staff's and board's views to be made public in the form of Public Information Notices, which are available at the IMF's Web site. Many countries also allow publication of the staff's report.

The IMF engages in multilateral surveillance to assess regional conditions and the world economy. The results of this work are published in the IMF's *World Economic Outlook* and its *Global Financial Stability Report*, each of which appears twice a year. The IMF has initiated a process of multilateral consultations with members in areas where there is a need for collective action, such as resolving global payments imbalances. However, the IMF can only operate through persuasion in these circumstances because it has no power over sovereign governments.

Technical Assistance The IMF provides technical assistance to its members, particularly lower-income members, on macroeconomic policies, financial sector regulation, and the collection and maintenance of databases. This assistance is financed by the IMF itself and the contributions of members as well as multilateral donors. The IMF integrates technical assistance with its surveillance and lending activities whenever possible. The IMF makes its assistance available through different mechanisms. It may send a team of economists to a country for a short-term project, or a resident advisor for a longer period. It also offers courses and training programs at the IMF Institute at its Washington, DC, headquarters and its six regional technical assistance centers and seven regional training institutes and programs in Africa, Asia, the Caribbean, Europe, the Pacific Islands, and South America.

Lending The IMF offers credit to countries with balance of payments imbalances; the institutional arrangements targeted for special circumstances are known as "facilities." The most commonly used mechanism is the Stand-By Arrangement (SBA), which is designed to assist countries with short-term problems, and usually lasts for 12 to 18 months. The Extended Fund Facility (EFF) was established to assist countries with longer-term problems and operates for up to three years. The Supplemental Reserve Facility (SRF) was introduced in 1997 to provide large amounts of funds to countries which experience sudden capital outflows, and these programs operate for two to three years. The Compensatory Financing Facility provides credit to countries subject to a shortfall in their export earnings or an increase in the cost of cereal imports. The fund also provides emergency assistance in special circumstances, such as natural disasters. The amount of money that a country can borrow from the fund its access limit depends on the type of program and the size of the country's quota. However, the IMF's lending can exceed this limit if a SRF is activated. All of these loans carry a market-based rate of charge, and large loans carry a surcharge.

The IMF also makes credit available to its poorest members through special facilities. The Poverty Re-

duction and Growth Facility (PRGF) is the successor to the Enhanced Structural Adjustment Facility (ESAF), which in turn followed the Structural Adjustment Facility (SAF). The new name was chosen to emphasize the twin goals of poverty reduction and growth.

A government that seeks external funding consults with the IMF about its needs and then signs a Letter of Intent (LOI), which is presented to the Executive Board for approval. A country can borrow up to 25 percent of its quota on liberal terms; larger amounts require activation of a lending arrangement. The LOI specifies the amount of available credit and also the policy conditions that the borrowing government agrees to implement. Once the LOI is signed, the IMF releases the credit in phased installments, which are tied to the adoption of the policies enumerated in the LOI. This practice is known as conditionality.

These conditions traditionally dealt with monetary and fiscal policies and a country's exchange rate. However, as upper-income countries turned to the private international capital markets in the 1980s, developing economies and subsequently the transition countries became the IMF's borrowers. The scope of conditionality expanded to include policies that were designed to improve the efficiency of these economies and promote growth. Structural policies included the liberalization of foreign trade, the restructuring of the financial sector, and privatization of state enterprises. Many of these conditions were consistent with the set of market-oriented policies that became known as the Washington consensus. By the 1990s, structural conditions appeared in virtually all of the IMF's programs.

The use of structural conditions, however, was widely criticized. The effectiveness of these conditions in promoting growth was questioned by critics, who pointed to counterexamples of successful development. China and India, for example, opened up their economies, but their governments played a role in the allocation of resources among firms and sectors. In addition, the record of implementation of structural conditions was mixed, in part because of domestic opposition.

The IMF reviewed its conditionality practices in 2002 and issued new guidelines. The Executive Board announced that the focus of structural conditionality in the future would be streamlined and limited to achieving the macroeconomic goals of the program. Moreover, the IMF acknowledged that programs were successful only when the borrowing government was committed to the process of reform a concept known as country ownership. Ownership required the full participation of the government in developing the program with a regard for the country's circumstances.

Criticisms The IMF has been the subject of criticisms from both wings of the political spectrum. As noted above, structural conditionality was often viewed as inappropriate and misguided. Some critics charged that it had unduly harsh impacts on the poor because of cutbacks in government spending on social policies. A number of nongovernmental organizations (NGOs) have protested the fund's use of conditions, and the IMF's Annual Meetings have been the subject of picketing and demonstrations.

The fund's macroeconomic conditionality was also criticized in the wake of the 1997 98 East Asian crisis, when Thailand, Indonesia, and Korea turned to the IMF for assistance. The IMF initially called for contractionary fiscal and monetary policies, consistent with its traditional view of a crisis as the consequence of expansionary domestic policies. However, these crises were driven by capital outflows that in some cases were prompted by contagion from neighboring countries, and the countries were already undergoing contraction as domestic private expenditures collapsed. The increases in interest rates recommended by the IMF for the purpose of supporting exchange rates also deepened the decline in output. The IMF subsequently revised these programs to allow fiscal deficits, but defended its call for higher interest rates as necessary to arrest further depreciations and inflation.

The IMF has also been charged with indirectly contributing to the emergence of these crises through its lending, an example of the phenomenon of moral hazard. Critics assert that foreign lenders have been willing to make risky loans because they believed that the IMF would provide financial support in the event of a crisis. Some of these critics have called for the abolition of the IMF, which they claim no longer serves the purpose for which it was created.

However, many observers believe that in a world of global capital flows and recurring financial crises, there is a need for an organization to deal with breakdowns in international monetary flows. Many of these collapses require collective action to resolve, and an institution like the IMF has the capability to coordinate the private and official responses.

However, in recent years the number of new lending arrangements has fallen sharply. Part of this decline is due to economic growth in developing countries and increased holdings of foreign exchange reserves by central banks. But many of the emerging markets can now obtain credit in the private capital markets and therefore do not borrow from the IMF except in crisis conditions. The IMF needs to redefine its role in a world of global financial markets if it wants to retain its relevancy.

See also asymmetric information; bail-ins; bailouts; balance sheet approach/effects; Bretton Woods system; capital flows to developing countries; currency crisis; financial crisis; global imbalances; gold standard, international; hot money and sudden stops; international financial architecture; International Monetary Fund conditionality; International Monetary Fund surveillance; lender of last resort; special drawing rights; Washington consensus; World Bank

FURTHER READING

Blustein, Paul. 2001. *The Chastening: Inside the Crisis that Rocked the Global Financial System and Humbled the IMF*. New York: Public Affairs.

———. 2005. *And the Money Kept Rolling in (and Out): Wall Street, the IMF, and the Bankrupting of Argentina*. New York: Public Affairs. An engaging account by a *Washington Post* reporter of the global financial crises of the 1990s and the subsequent one in Argentina and the IMF's responses.

Boughton, James M. 2001. *Silent Revolution: The International Monetary Fund, 1979 1989*. Washington, DC: IMF. A history of the IMF from the IMF's own historian.

International Monetary Fund, Independent Evaluation Office. 2003. *IMF and Recent Capital Account Crises: Indonesia, Korea, Brazil.* Washington, DC: IMF.

———. 2004. *Report on the Evaluation of the Role of the IMF in Argentina, 1991 2001.* Washington, DC: IMF. Thorough analyses of the fund's involvement in some of the major crises by the IMF's own Independent Eva luation Office.

James, Harold. 1996. *International Monetary Cooperation since Bretton Woods.* Washington, DC: IMF and Oxford University Press. A history of the IMF.

Stiglitz, Joseph E. 2002. *Globalization and its Discontents.* New York: W. W. Norton. A far ranging examination and critique of the IMF's policies.

Woods, Ngaire. 2006. *The Globalizers: The IMF, the World Bank, and Their Borrowers.* Ithaca, NY: Cornell Uni versity Press. An analysis of the relations of the IMF and World Bank with their member governments and their programs in Mexico, Russia, and Africa

JOSEPH P. JOYCE

■ International Monetary Fund conditionality

The International Monetary Fund (IMF) lends to its member nations with balance of payments problems in order to promote international financial stability. The government of a country that borrows from the IMF agrees to implement economic policies to rec tify the situation, and the IMF disburses the credit as these policies are implemented. This monitoring mechanism is known as conditionality. The IMF uses conditionality to ensure that the borrowing country is resolving its problems and that the IMF's financial resources will be available to other mem bers. The nature of conditionality has evolved over time in response to changes in the role of the IMF in the international economy and with its members.

Lending Procedures A country can borrow an amount equal to 25 percent of its quota with the IMF on liberal terms; larger amounts require the estab lishment of a policy program. The IMF uses a number of administrative programs to make credit available to its members. The most commonly used are the Stand-By Arrangement (SBA), which is de signed to assist countries with short-term (1 to 2 years) problems, and the Extended Fund Facility (EFF) for countries with longer-term (typically 3 years) problems. Both programs carry a market-related interest rate. Between 1953 and 2007, there were 835 SBAs approved and 76 EFFs.

The IMF also lends to its poorest members at a below-market rate through the Poverty Reduc tion and Growth Facility (PRGF), which is the successor to the Structural Adjustment Facility (SAF) and Enhanced Structural Adjustment Facility (ESAF). The interest rate on a PRGF loan is 0.5 percent, and the lending typically takes place over a 3-year period. There are approximately 75 mem bers of the IMF are eligible for the PRGF. There were 199 SAF/ESAF/PRGFs approved between 1953 and 2007.

After consultation with the IMF, a government that seeks credit signs a Letter of Intent, which spe cifies the amount of credit to be made available and the policy conditions that the government agrees to implement. A country applying for a PRGF must submit a Poverty Reduction Strategy Paper, which describes the steps that its government will undertake to foster growth and reduce poverty. Once the ar rangement is approved by the IMF's Executive Board, the IMF releases the credit in phased install ments, which are tied to the adoption of the enu merated policies.

The IMF monitors compliance with the policy conditions through a variety of institutional mecha nisms. Prior actions are prerequisites that a country undertakes before the program actually begins. Per formance criteria are specific conditions that a country must meet before it receives a disbursement of credit. Quantitative performance criteria include macroeconomic policy targets, such as monetary aggregates and fiscal balances. Structural perfor mance criteria deal with an economy's efficient use of its resources and include measures such as deregula tion. There may also be structural benchmarks that provide additional information about a country's program implementation.

The IMF will suspend a program if its conditions are not met. Waivers can be granted, however, when implementation has been delayed due to circumstances outside a government's control, such as lower than expected tax revenues or insufficient external financing. The IMF's Executive Board can perform a program review in situations where the original program requires revision due to new developments.

Analytical Tools The IMF draws on analytical frameworks in formulating the policies it recommends to borrowing countries. Among these is financial programming, which was developed at the IMF in the 1950s. This model links balance of payments deficits under fixed exchange rates to domestic monetary policy. It is based on two behavioral equations (the demand for money and import demand), two identities (the base money supply and the balance of payments), and an equilibrium condition (money market equilibrium). Private capital flows were considered to be fixed, as they often were in the Bretton Woods era. Financial programming is based on data that are usually available to policymakers, and the model's simplicity was an asset in the postwar era when data limitations were more common.

In the model, a rise in a central bank's holdings of domestic credit leads to an increase in national income and consequently imports. The immediate result is a current account deficit and a loss of reserves. In the long run, the initial change in domestic credit is completely offset by the change in foreign reserves. The solution to a balance of payments deficit lies in curbing the growth of domestic credit. This adjustment in monetary policy can be linked to the government's fiscal position, since in many developing countries a fiscal deficit is financed by the central bank.

The other model developed at the IMF is the absorption model, which relates the state of the current account to domestic expenditures and output. In this model, the current account reflects the difference between domestic output and absorption, which is the domestic demand for goods and services. A surplus is recorded when national output is higher than absorption, and a deficit in the opposite situation. Contractionary spending policies are required to lower or eliminate the deficit.

The policies in IMF-supported programs have been based on these models. Fund programs usually call for a tightening of monetary and fiscal policies in order to reduce total expenditures, and the IMF's performance criteria often include limits on domestic credit expansion and government expenditures and increases in tax revenues. In addition, the IMF in some cases has recommended a depreciation of the exchange rate in order to switch expenditures to domestically produced goods and stimulate exports.

There have been a number of criticisms of these models. The demand for money is often not stable and predictable, and changes in velocity affect the transmission of monetary changes within the economy. Moreover, the liberalization of the capital account and the increase in private international capital flows have introduced new sources of volatility into the financial sector.

The IMF's macroeconomic policy conditionality came under sharp criticism during the 1997–98 East Asian financial crisis. The cause of the crisis was a rapid outflow of capital, not expansionary government policies. Critics claimed, however, that the IMF misjudged the nature of the crisis and promoted contractionary policies at a time when output in the crisis countries was already falling. As the crisis progressed, the IMF realized that the emerging fiscal deficits reflected the crisis-induced declines in output and relaxed its call for fiscal consolidation. The IMF defended the use of higher interest rates, however, as necessary to prevent further depreciation of the exchange rates and an outbreak of inflation.

The IMF developed new tools in response to the emergence of capital account crises, such as those that occurred in East Asia and in Mexico in 1994–95. The balance sheet approach specifies the assets and liabilities of the different sectors of the economy, including the financial liabilities that are denominated in a foreign currency. These represent a source of vulnerability if foreign creditors lose confidence in domestic borrowers' ability to fulfill their obligations and decide to withdraw their funds. Economists at the IMF have also sought to develop indicators of

impending currency crises that would serve as early warning systems and give the IMF and the country at risk time to prevent the crisis.

Structural Conditionality During the 1980s, the IMF's lending focused on developing economies drawn into the debt crisis. In the following decade, the transition economies that emerged after the dissolution of the Soviet Union also became fund borrowers. The governments of these countries were concerned about economic growth as well as external stability and sought to enact policies that would accelerate growth. In response, the IMF and its sister institution the World Bank devised structural adjustment policies to improve the efficiency of an economy and promote growth. These policies originally were included in the IMF's concessionary lending programs for the poorest members, but over time appeared in the SBAs and EFFs as well.

Structural adjustment policies can include fiscal reform, the privatization of government-owned enterprises, and the reorganization of the financial sector. These measures are intended to replace government allocation of resources with those of the market. Some of these reforms appeared in the set of market-oriented policies known as the Washington consensus, which encapsulated the state of thinking among many policymakers and advisors in the late 1980s and early 1990s regarding the steps needed to advance growth in developing countries.

Over time the scope and number of conditions rose. The specification of conditions became more detailed because implementation of structural policies is more difficult to assess than the traditional macroeconomic conditions. The IMF (2001b) reported that more slippage took place in the implementation of structural conditions than occurred with conditions in the traditional area of macroeconomic policies.

Structural conditionality became a focus of criticism. First, many of these policy areas were outside the IMF's expertise in macroeconomics. Second, there was limited evidence connecting many of the structural conditions to increased growth. Countries that had established strong records of growth in the past, such as Taiwan and Korea, had allowed domestic firms to develop before opening up their economies. Third, in many cases there was limited domestic political support for structural conditions, which were viewed as an infringement of national sovereignty.

The criticisms of structural conditionality peaked after the East Asian crisis. Goldstein (2003) reported that the number of structural conditions totaled about 140 in the case of Indonesia, 90 in Korea, and 70 in Thailand. In the case of Indonesia, many of these conditions dealt with issues that had only tangential relevance to the goal of balance of payments equilibrium. Moreover, critics charged that the IMF itself had indirectly precipitated the crisis through its earlier promotion of financial liberalization.

The IMF undertook a review of its conditionality practices in response to the criticisms and issued new guidelines in 2002, which were reviewed in 2005. These guidelines emphasize the need for national ownership of the policies contained in a lending program. The IMF has subsequently sought to deepen government involvement in the formulation of programs. The Poverty Reduction Strategy Papers for the PRGF programs, for example, are prepared by governments in consultation with civil society and other participants. The new guidelines also called for the streamlining of conditions to those areas that were seen as essential to reestablishing a sustainable balance of payments.

Compliance and Recurring Usage Other aspects of IMF conditionality that have received scrutiny and discussion are the implementation and the recurring use of the IMF's resources. In some cases governments do not implement all the conditions contained in their Letter of Intent, and in response the IMF suspends its disbursement of funds. Countries with interrupted programs do not establish strong records of economic performance. Empirical analyses of these situations reveal that domestic political economy factors, such as a lack of cohesion within a government, can impede the progress of a program.

Governments that do not complete one program may adopt another. A study by the IMF's Independent Evaluation Office (2002) found that when prolonged usage is defined as the adoption of

IMF programs for 7 or more years in a 10-year period, then 44 countries were prolonged users between 1971 and 2000. The report also found that prolonged usage expanded consistently since the 1970s, and the authors attributed this development in part to the expansion of the goals of the IMF's lending programs. Studies of frequent usage find that these recurring borrowers do not achieve a stage of economic performance in which they can dispense with the IMF's assistance. Incomplete implementation and frequent usage detract from the effectiveness of fund programs and may delay the establishment of national policies and institutions to resolve continuing deficits.

See also balance sheet approach/effects; Bretton Woods system; capital flows to developing countries; currency crisis; early warning systems; expenditure changing and expenditure switching; financial crisis; financial liberalization; hot money and sudden stops; International Monetary Fund (IMF); International Monetary Fund surveillance; money supply; Washington consensus

FURTHER READING

Bird, Graham, Mumtaz Hussain, and Joseph P. Joyce. 2004. "Many Happy Returns? Recidivism and the IMF." *Journal of International Money and Finance* 23 (2): 231 51. A study of recurring lending arrangements.

Corden, W. Max. 2001. "The World Financial Crisis: Are the IMF Prescriptions Right?" In *The Political Economy of International Financial Crisis*, edited by Shale Hor owitz and Uk Heo. Lanham, MD: Rowman and Lit tlefield, 4 61. A comprehensive review of the features and impact of fund conditionality.

Drazen, Allan. 2002. "Conditionality and Ownership in IMF Lending: A Political Economy Approach." *IMF Staff Papers* 49 (Special Issue): 36 67. An analysis of the roles of conditionality and ownership within IMF pro grams.

Easterly, William. 2005, "What Did Structural Adjustment Adjust? The Association of Policies and Growth with Repeated IMF and World Bank Adjustment Loans." *Journal of Development Economics* 76 (1): 1 22. An empirical study of recurring lending arrangements.

Ghosh, Atish, Charis Christofides, Jun Kim, Laura Papi, Uma Ramakrishnan, Alun Thomas, and Juan Zaluendo. 2005. *The Design of IMF Supported Programs*. IMF Occasional Paper No. 241. Washington, DC: IMF. A review of the objectives and development of IMF pro grams from 1995 to 2000.

Goldstein, Morris. 2003. "IMF Structural Programs." In *Economic and Financial Crises in Emerging Market Economies*, edited by Martin Feldstein. Chicago: Uni versity of Chicago, 363 437. A comprehensive review of the features and impact of Fund conditionality.

International Monetary Fund. 1987. *Theoretical Aspects of the Design of Fund Supported Adjustment Programs*. Washington, DC: IMF. The IMF's explanation of the analytical frameworks underlying conditionality.

———. 2001a. *Conditionality in Fund Supported Pro grams Policy Issues*. Washington, DC: IMF. A review of the objectives and history of conditionality.

———. 2001b. *Structural Conditionality in Fund Supported Programs*. Washington, DC: IMF. An assess ment of the record of structural conditionality.

———, Independent Evaluation Office. 2002. *Evaluation of Prolonged Use of IMF Resources*. Washington, DC: IMF. A study of recurring lending arrangements.

Joyce, Joseph P. 2006. "Promises Made, Promises Broken: A Model of IMF Program Implementation." *Economics and Politics* 18 (3): 339 65. An examination of the im plementation of program conditions.

Khan, Mohsin S., and Sunil Sharma. 2003. "IMF Con ditionality and Country Ownership of Adjustment Programs." *World Bank Research Observer* 18 (2): 227 48. An analysis of the roles of conditionality and own ership within IMF programs.

JOSEPH P. JOYCE

■ International Monetary Fund surveillance

Surveillance, as practiced by the International Monetary Fund (IMF), is the process of overseeing the functioning of the international monetary system and certain economic and financial policies of IMF member states. The legal mandate for this activity is Article IV of the IMF Articles of Agreement, which was rewritten after the 1973 collapse of the Bretton Woods system, when exchange rates were no longer

anchored by gold, and currency values of most of the major industrial countries floated. Section 1 of the new Article IV (adopted in 1978) requires "each member . . . to collaborate with the Fund and other members to assure orderly exchange arrangements and to promote a stable system of exchange rates." Section 3 requires the IMF to "oversee the international monetary system in order to ensure its effective operation, and [to] oversee the compliance of each member with its obligations under Section 1 of this Article." To that end, the fund is required to exercise "firm surveillance" over member countries' exchange rate policies, and those members are required to consult with the fund on their policies.

Practice The two main methods of surveillance are periodic (usually annual) consultations between the fund and each member and the preparation by the fund of periodic reports on global economic and financial conditions. The first method   the Article IV consultation   is known as bilateral surveillance, and the second as multilateral surveillance. Since 2006, the second form has been expanded to include multilateral consultations, in which the IMF convenes discussions involving several members and covering issues of mutual interest and systemic consequence.

An Article IV consultation is primarily a confidential discussion between the fund and the country's monetary authorities (the finance ministry and the central bank). Increasingly in recent years, these consultations have also provided a means for the IMF to convey its views to the general public and to financial markets. Each consultation begins with a visit to the country by a team of IMF staff, referred to as a staff mission. The mission meets with officials in the finance ministry, the central bank, other official agencies as needed, and usually with representatives of the business community, labor, and civil society. At the conclusion of these talks, the mission chief presents a preliminary report to the government setting out the fund staff's assessment of economic and financial conditions in the country and its advice on exchange rate and related policies. With the government's permission, he or she may also hold a

press conference to make the mission's findings known more widely.

After returning to the IMF headquarters in Washington, DC, the mission prepares a more detailed report, which the IMF Executive Board discusses. The board chairman then issues a summary of the discussion that sets out the fund's policy advice in final form. With the government's consent, the fund may then issue a public information notice (PIN) based on the summing up. Each PIN is published on the IMF's Web site, often together with the staff report.

The core of multilateral surveillance is the *World Economic Outlook* (WEO) exercise, which is normally conducted semiannually. The staff prepares a set of studies on world and regional economic developments, including forecasts and medium-term scenarios under specified assumptions and detailed analyses of key policy issues. The WEO reports have been published since 1980 and have become the flagship publication of the IMF. The other major component of multilateral surveillance is the preparation of the semiannual *Global Financial Stability Report*, which includes an analysis of developments in financial markets, incorporating both official and private capital flows.

Limitations Although the IMF considers surveillance to be its core activity and responsibility, critics argue that bilateral surveillance is an ineffective way to influence and discipline countries' policies. In broad terms, the success of surveillance depends on the quality of the fund's analysis and policy advice, the effectiveness of the dialogue and the communication of the fund's views to the authorities and to the public, and the willingness of members to accept the fund's advice. Although specific criticisms can always be leveled at specific aspects of the dialogue and the conclusions regarding a country's economic and financial policies, the central difficulty in bilateral surveillance is in persuading the authorities to act on the basis of advice from the IMF. In many cases, governments may have shorter-term domestic objectives or domestic political constraints that conflict with the longer-term and more international objec-

tives promoted by the IMF. In other cases, there may be genuine differences of economic philosophy between the fund and the authorities.

The voluntary nature of the response to IMF surveillance contrasts with the position of countries that borrow from the IMF. For the latter, because the fund's lending is generally conditional on the borrower's adherence to an agreed program of policy improvements, a country that needs the fund's financial support faces practical constraints in its choice of policies. Advanced economies with ready access to private capital markets and developing countries that have accumulated foreign exchange reserves by running external payments surpluses are thus freer to accept or reject policy advice from the IMF than are countries with more limited financial means. These differences, of course, reflect disparities in the world economic order, not an institutional weakness in surveillance.

History When the IMF began operations in 1946, it conducted consultations only with countries that availed themselves of the temporary provisions of Article XIV of the Articles of Agreement. Under the terms of that article, these countries were still imposing exchange restrictions on foreign trade and other current account flows, and were obligated to consult with the IMF on their plans to eliminate those restrictions when practicable. At the beginning, most members maintained such restrictions, and the number fell only gradually. By 1960, only 10 of the fund's 68 members had accepted the obligations of Article VIII prohibiting further recourse to current-account restrictions. In 1961, however, the number of Article VIII countries more than doubled, in response to the restoration of current account currency convertibility across much of Europe. Thus the need for Article XIV consultations began to fall.

In that same year, 1961, the United States   the IMF's largest member country and the first to accept the obligations of Article VIII in 1946   agreed voluntarily to hold annual consultations with the fund "as a means of exchanging views on monetary and financial developments." From that point on, an increasing number of countries volunteered to hold these "Article VIII" consultations. None of this activity was yet called "surveillance," but it did lay the groundwork for the more formal and obligatory system of consultations that would come later.

IMF surveillance, in the sense that the term is used today, began with the collapse of talks by a committee of the IMF's governing board in 1974. For two years, this Committee of Twenty tried to devise a new system of stable exchange rates to replace the Bretton Woods par-value system that had been abandoned in 1973 after the United States terminated the gold convertibility of the dollar. When it finally became clear that the United States and some other major countries were not prepared to peg their exchange rates again, the committee fell back on the subtle distinction of promoting a "stable system of exchange rates" rather than a "system of stable exchange rates."

In order to achieve stability while exchange rates were floating, the Committee of Twenty reasoned that the IMF could play a key role by overseeing the international financial system and the macroeconomic policies that underpinned it. The fund's charter was then amended in 1978 to let countries choose their own exchange arrangements   pegged to another currency or a basket of currencies, freely floating, or managed more or less flexibly   subject to a requirement to consult regularly with the IMF on how they were implementing their policies. As noted above, the IMF was given a mandate to exercise "firm surveillance" over the conduct of members' exchange rate policies and to oversee the functioning of the system. Hence, modern surveillance was born.

A tug-of-war quickly emerged between the desirability of focusing on key issues and problem cases, on the one hand, and trying to be comprehensive and even-handed with all member countries, on the other. If the IMF conducted annual consultations with all members covering all policies and conditions with significant effects on exchange rates and international payments balances, staff resources would be stretched thin and the process would become routine. If it was more selective, it risked missing some brewing problems and violating the requirement that

the fund treat all of its members in an even-handed manner. In either case, surveillance would fall into disrepute and again become ineffective. Attempts were made to hold ad hoc special consultations with countries that were suspected of manipulating their exchange rates for competitive advantage, but resistance to such procedures proved to be strong. Regularly scheduled bilateral consultations continued to be the primary means of communicating the fund's advice to members.

In two respects, the effectiveness of surveillance was enhanced during the 1990s. First, the Interim Committee (the successor to the Committee of Twenty) agreed in general terms on standards for the conduct of macroeconomic policies, most notably through the declaration of the Partnership for Sustainable Global Growth in 1996. Those standards, which the IMF managing director dubbed the "Eleven Commandments," enabled the fund to give more pointed advice to countries whose policies appeared to be out of line. Second, the fund adopted new procedures to open up the dialogue to wider participation and to allow the publication of staff reports and PINs. This new transparency facilitated public scrutiny and international comparisons of countries' policies and made it more difficult for governments to ignore the fund's recommendations.

By the late 1990s, two other problems with the effectiveness of surveillance had become clear. First, the quality and timeliness of financial data published (or even provided privately to the fund) by monetary authorities were not uniformly high. In a few cases, countries that had been particularly reluctant to reveal the state of their official reserve position were hit by severe financial crises. In response, the IMF embarked on a major effort to encourage members to improve financial reporting, including the establishment of standards for data dissemination and a system for disseminating data electronically to the public. Second, the fund's traditional emphasis on dialogue with members' monetary authorities occasionally left it with insufficient information about developments in financial markets. This problem was particularly acute in so-called emerging market countries, where bank loans and other private capital

flows were major sources of external financing. After a wave of financial crises hit emerging market countries in the late 1990s, the IMF responded by greatly expanding its contacts with and analysis of capital markets and strengthening its capacity to assess countries' vulnerability to crises.

In 2006, when the world economy was growing rapidly but was at risk because of large and growing imbalances among the large economies, it again appeared that the tools of IMF surveillance needed to be further strengthened. That led to the inauguration of multilateral consultations, the first of which was aimed at enhancing the dialogue among the countries with the largest surpluses and deficits and seeking a commonly acceptable solution.

See also balance of payments; Bretton Woods system; currency crisis; exchange rate regimes; global imbalances; gold standard, international; international financial architecture; International Monetary Fund (IMF); International Monetary Fund conditionality; international policy coordination; twin deficits

FURTHER READING

The following sources provide histories of the evolution and operations of the IMF:

Boughton, James M. 2001. *Silent Revolution: The International Monetary Fund 1979 1989*. Washington, DC: International Monetary Fund.

de Vries, Margaret Garritsen. 1976. *The International Monetary Fund 1966 1971: The System under Stress*. 2 vols. Washington, DC: International Monetary Fund.
    . 1985. *The International Monetary Fund 1972 78: Cooperation on Trial*. 3 vols. Washington, DC: International Monetary Fund.

Horsefield, J. Keith. 1969. *The International Monetary Fund, 1945 1965*. 3 vols. Washington, DC: International Monetary Fund.

International Monetary Fund. Available at http://www.imf.org. The best source for up to date information on IMF surveillance.

For specific analyses of surveillance, the following sources are useful:

International Monetary Fund, Independent Evaluation Office. 2006. *Multilateral Surveillance*. Washington, DC: International Monetary Fund.

James, Harold. 1995. "The Historical Development of the Principle of Surveillance." *IMF Staff Papers* 42 (December): 762 91.

———. 1996. *International Monetary Cooperation since Bretton Woods*. Washington, DC: International Monetary Fund and Oxford University Press. A one volume history of the IMF's role in the financial system, including the context in which surveillance has been practiced.

Pauly, Louis. 1997. *Who Elected the Bankers? Surveillance and Control in the World Economy*. Ithaca, NY: Cornell University Press.

JAMES M. BOUGHTON

■ international policy coordination

The coordination of economic policies among countries is a means to promote economic and financial stability and a way to orchestrate growth, particularly in periods of depression, without the costs of additional inflation. At first it was viewed in terms of coordinating monetary policies to reduce the tendency for interest rates to rise and fluctuate as countries attempted to emerge from recession (the Bonn Summit, 1978), or to eliminate the need for competitive devaluations as economies attempted to expand in a world of increasing trade flows and capital links (European Monetary System, 1979). Later, as inflation performance came into line with the targets set for it, the focus shifted to fiscal coordination (the Plaza Agreement, 1985), joint exchange rate targeting (Louvre Accord, 1987), and to institutional coordination within an economy or group of economies. In the early 21st century, policy coordination is mostly implicit: often discussed, but seldom in the form of explicit rules or agreements. What survives is the staff work behind the annual Group of Eight (G8) summits of the world's largest economies.

Rationale for Coordination The rationale for coordinating the policies of interdependent, but politically sovereign, policymakers is usually efficiency. Formal game theory models show that policy bargains can always be found that will leave some countries better off without others being worse off in terms of their own policy objectives. By cooperating to improve the performance of domestic policy targets and to account properly for the consequences of the spillover effects of their own policies on others, each country may achieve its own objectives, relative to what could be obtained by adopting the alternative strategy of determining its policies independently. This argument is well established in theory and has been demonstrated formally in a range of analytic and empirical models (Hamada 1976; Canzoneri and Gray 1985; Hughes Hallett 1986).

It is sometimes useful to think in terms of absolute and relative coordination. Absolute coordination focuses on the overall stance of policy to ensure that countries do not, in pursuing their own interests, adopt policies that are too tight or too loose for the world as a whole. This has to do with efficiency, or comparative advantage in getting better outcomes for all. Relative coordination, meanwhile, is concerned with adverse spillovers, redistribution, and the relative positions of countries. It therefore tends to focus on exchange rates, trade imbalances, or fiscal imbalances. Policymakers usually appear to be more interested in relative coordination (for example, the current trade imbalances), but turn to issues of absolute coordination in periods of widespread recession or inflation. Given that, the academic literature has been concerned with the following issues:

1. How large are the gains from policy coordination in practice, and how will they be distributed among participating economies?

2. What are the key determinants of the size and distribution of the payoffs?

3. Are the policies of a coordinated policy bargain stable? The problem is that once a country is persuaded to adopt policies different from those that are optimal in a noncooperative context, the other countries have an incentive to reoptimize once again on the basis that their partners are now committed to particular policy paths.

4. What are the obstacles to coordination? There may be political and institutional

reasons why it is more difficult to secure an agreement on certain policy variables, but easier on others.

5. To what extent are coordinated policies sensitive to errors, or to disagreements among policymakers about the assumed policy responses in the economy or what the appropriate policy priorities might be?

Levels of Policy Cooperation and Coordination
The potential benefits and costs of policy coordination will depend on the degree of cooperation among policymakers. There is a natural hierarchy here:

Information exchanges. Countries freely exchange information about their targets, priorities, information sets, and how they think those targets would respond to domestic or foreign policy changes. But given that information, national policymakers would continue to make decisions in a decentralized, autonomous way. By pooling their information, countries could make gains by eliminating incomplete or faulty information over objectives, expectations, and assumed policy responses.

Crisis management. Coordination would respond to episodes of particular difficulty in the international economy, involving policy changes that are particular to that episode. It might involve ad hoc policy adjustments for difficult periods, when imbalances and current policies have interacted to yield a crisis.

Avoiding conflicts over shared targets. Shared targets arise where countries actually target the same variable (for example, a mutual exchange rate), or where they target variables that are linked by an identity that cannot be relaxed via policy (for example, the "n 1 problem" in a set of current accounts, which will be discussed later). In this case, coordination could generate gains by means of agreements that prevent countries from setting incompatible targets for the same variable, or by preventing countries from attempting competitive policy changes that cannot all be achieved simultaneously.

Intermediate targeting. A limited degree of coordination may be achieved when countries jointly control the variables that form the main monetary, fiscal, or financial links between their economies. In that case, the linked variables are treated as intermediate targets, being instrumental in obtaining better results for the other targets but without significance in themselves. Intermediate targets may or may not be shared targets, although the scope for fruitless competition over shared targets makes them the most obvious candidates.

Partial coordination. Countries cooperate in achieving certain targets but may aim at other targets uncooperatively or according to some preassigned (national) rules. It is often suggested that countries should coordinate their monetary policy, leaving fiscal policy (which is hard to manipulate in the short term) for domestic targets as happens in the European Monetary Union. Fiscal policy needs to be sustainable, however, so a minimal degree of fiscal coordination may be required to permit effective monetary coordination.

Full coordination. Countries adopt a certain bargain across all targets and fiscal, monetary, or exchange rate instruments. They would aim to maximize the gains over the noncooperative policy settings, subject to an acceptable distribution of those gains among participants or countries.

The Gains from Coordination Studies designed to evaluate the gains from policy coordination empirically have found the benefits to be valuable but small. The gains for industrial countries from cooperating are generally thought to be in the range of 0.5 1.5 percent of gross national product (GNP) to each country, over the best available noncooperative outcomes (Oudiz and Sachs 1984; Hughes Hallett 1986). Nevertheless, these gains turn out to increase significantly with the size and persistence of external shocks and the reputations (for consistency) of the governments concerned (Currie, Levine, and Vidalis 1987). It can be debated whether one-half of 1 percent of GNP is only a "small" gain. It would represent a significant amount of extra productive capacity if fully invested. But it is not large compared to annual growth rates. Moreover, if the expected gains are small relative to the imprecision with which policies can be implemented, those gains may be hard to realize in practice. Consequently, the robustness of these gains is particularly important.

Far less attention has been given to the likely distribution of the gains among countries. Empirical studies generally find that the gains from cooperation are asymmetric, and it is usually very difficult to find ways of improving the lot of the countries that benefit least from coordination (Hughes Hallett, Holtham, and Hutson 1989). These are important results because they suggest that, whatever the overall gains, (1) it will be hard to secure *and maintain* a coordination agreement in the face of significant uncertainties, and (2) if those who make the gains and those who shoulder the burden of adjustment are different sets of people, then securing any agreement will be politically difficult.

Cooperation through Information Exchanges The gains from coordination relative to noncoordination may well be smaller than those of efficient noncooperative policies relative to strategies that ignore predictable policy changes abroad or elsewhere in the system. Thus coordination in the sense of *information exchanges,* rather than a negotiated solution across all variables, may supply the major part of the improvements. If this is so, an important function of international policy discussions or policy meetings is going to be the exchange of information among policymakers concerning their policies, their aims, and the state of their economies. The implication is that much of the value of policy coordination may in fact lie in tying the national policymakers into a regime in which it is far more difficult for them to make poorly designed or misguided policy choices either through misplaced self-interest or through ignorance rather than in introducing qualitatively different policies.

Coordination Substitutes: Avoiding Conflicts over Shared or Linked Targets None of the original studies of policy coordination considered exchange rates to be a target of policy. Yet the policy debate has often been concerned with exchange rate management, with the aim of either stabilizing exchange rates or of making controlled realignments. When exchange rates are included among the targets during policy selection, they can either be included in the associated objective function evaluations (in which case exchange rate stability becomes a target

in its own right) or excluded from those evaluations (in which case exchange rates are just an intermediate target, instrumental in securing improvements elsewhere, but of no interest in itself). In the former case, the gains from coordination appear larger than before, at about 3 6 percent of GNP as estimated across seven multicountry models. If exchange rates are treated merely as intermediate targets, the gains are significantly smaller and much as in the earlier literature cited above (Holtham and Hughes Hallett 1987).

The point here is that an exchange rate is a shared variable whose domestic impact is the same whether exchange rate changes originate at home or abroad. By contrast, most other variables exert impacts that are significantly smaller internationally than domestically. The need to limit exchange rate spillovers is therefore greater in problems with shared targets than in those with other linkage variables.

Second, success with shared targets, whether treated as intermediate or not, requires a measure of coordination in how they are controlled and also some consensus about the target path that they should pursue. If this is missing, countries will inevitably waste policy power pushing against one another in a vain attempt to achieve the impossible. This holds in particular for the n 1 problem. In that case, there is one fewer independent exchange rate (or current account) than the number of independent policymakers so that not all n target values can be reached at once. At least one must be missed, or be allowed to float free. On the other hand, it is not true that any agreed target path is better than none. Jointly specifying exchange rate target paths independently of any other objectives is frequently more damaging than moderate disagreement about what that target path should be. Hence the problem appears to be one of choosing an appropriate set of target paths, not one of securing precise agreement on the policies to support some path.

In addition, not only may intermediate targets serve to promote relative coordination between economies, they may also be conducive to absolute coordination. This is an important part of the rationale for joint monetary targeting or for a system

of exchange rate targets. Nevertheless there are important qualifications to this argument. The gains to exchange rate targeting may be small, hard to achieve, or badly distributed across countries. It is not just that some countries do better than others. Some may actually lose compared with the no-targeting case. Moreover, the target paths have to be very carefully chosen to achieve any of these gains (Hughes Hallett 1992). This suggests that the losses attributable to attempts to manipulate monetary policies (or exchange rates) within a given regime are generally small but may become significant in particular episodes when countries have uniform objectives or attach great weight to one objective. At other times, the potential gains to eliminating competitive policy-making are easily outweighed by the losses resulting from the attempt to achieve an extra, and possibly inappropriate, exchange rate target.

In short, since any targeting regime implies an additional constraint, the sacrifice in the other targets, which this extra constraint implies, must be less than the efficiency gains from a more effective design of the policies or a more appropriate allocation of interventions between national policy instruments and policy institutions.

The Dangers of Coordination Because the gains may not be large, obstacles to coordination may easily create losses that overwhelm those potential gains. Policymakers may decide not to cooperate if they think that other countries will not stick to their part of the bargain, or if they think that there are potential information errors or other shocks that would invalidate their calculations, or if they think that they may have misestimated the priorities of other players, or if there is a serious possibility that policymakers are uncertain or differ in their views of how the economy works (Holtham and Hughes Hallett 1992).

Sustainability Will policymakers stick to their part of the bargain? Or will policymakers cheat by redesigning their future policies once they have maneuvered the other participants into taking certain actions? The first question involves breaking a bargain, with one party finding it advantageous to deviate from what was previously agreed once the other

participants have established their policies. In that case governments try to cheat on one another. The second is the usual time-inconsistency problem in which governments cheat on the private sector over time. The latter does not involve a bargain as such but represents an internal inconsistency in one of the policy sequences. A standard example is discussed by the economist Rogoff (1985), who points out that policy coordination may actually be welfare-decreasing if the coordination process eases the constraints imposed on governments engaging in inflationary expansions. The reason is that, in reality, there is more than one set of actors per economy (the government, the central bank, and a private sector whose forward-looking expectations affect the exchange rate and, hence, policy success) so that coordination among governments may represent a coalition against the private sector or central bank, rather than full coordination. Although the example is specific, it is a general point that coordination without credibility may be counterproductive. Nevertheless, coordinated policies will prove robust against cheating if the losses the first country (or actor) would suffer under retaliation (or preemptive cheating) by its opponent are larger than the gains that the same country (or actor) could make by unopposed cheating.

Policy Errors Shocks to the system can affect the incentive to coordinate. But those shocks have to be both large and persistent (Canzoneri and Minford 1986). By contrast, coordination may reduce policy errors because decision-makers try to share the risks rather than offload them onto their rivals. Performance indicators are generally aggregations of national targets that are affected in different and mutually conflicting ways by external shocks, while coordinated policies follow by aggregating those national indicators into some global performance index. Coordinated policies will, therefore, normally be *more* robust than noncooperative policies, which lack that extra aggregation and hence that extra ability to cancel out errors.

Interinstitutional and Rule-Based Coordination Another form of policy coordination is rule-based coordination. In this case benefits arise from

the credibility associated with committing to a set of rules that are well designed and capable of being monitored easily. Rule-based coordination has the advantage that it may be used to avoid the inefficiencies of short-term policymaking, or where there are temptations to manipulate or opportunistically free ride on the expectations of others. One example is the "fiscal leadership" model in which fiscal policymakers are made to decide their policies first, with an eye to their long-run commitment to public services and the stability of public finances, under the threat that an independently run monetary policy that follows can take away any short-term or special interest gains that the fiscal leaders may be tempted to create for themselves (Hughes Hallett and Weymark 2007).

In fact, leadership under suitable conditions (that the players' reaction functions form an acute angle so that their policies could be aligned) will always induce a degree of coordination and the gains will increase the more the policymakers differ in terms of goals and priorities. A simple way to implement this idea is to give fiscal policymakers a debt target. Being a stock, not a flow, such a rule would induce the leadership we require automatically. More generally, this option opens up the whole area of interinstitutional coordination. The early studies ignored that issue, but more recent work has focused on the interactions among institutions (Foreman-Peck et al. 2007; Pappa 2004). Policy coordination is often more important in this context because the effects go through the domestic markets as well as through the international or foreign markets. Interinstitutional coordination is therefore likely to have a greater effect than intercountry coordination.

Explicit rules incorporating an automatic degree of coordination may also act as an important discipline on the actions of governments in the international arena. Although democratic governments should, and do, have appreciable freedom of action, the public choice literature emphasizes that pressure groups and special interests may exert undue influence on government policy, leading to inefficient outcomes for the economy as a whole. There may be scope for constraining government action to

more efficient outcomes on the basis that "what benefits others benefits everyone." Such rules may be internal, as in the separation of fiscal and monetary powers in some countries, or externally imposed as in exchange rate targeting or common inflation policies.

Going Forward Academic research has produced an extensive literature on policy coordination. Much of the work has been theoretical and based on simplified models. As these have abstracted from some of the difficulties of which policymakers are most aware, the literature may have a flavor of unreality to practitioners. Moreover, much of the impetus to cooperation has come from experience, not from explicit demonstrations of the benefits. Bringing theory to bear on current concerns is therefore unfinished business. But there is little sign as yet that our understanding of policy coordination needs to be revised (Canzoneri et al. 2005).

See also Bonn Summit; contagion; discipline; euro; European Monetary Union; exchange rate regimes; foreign exchange intervention; Louvre Accord; Maastricht Treaty; monetary policy rules; Plaza Accord; Smithsonian Agreement; spillovers; time inconsistency problem

FURTHER READING

Bryant, Ralph C. 1995. "International Coordination of National Stabilization Policies." Brookings Institution, Washington, DC. A background framework.

Canzoneri, Matthew, Robert Cumby, and Behzad Diba. 2005. "The Need for International Policy Coordination: What's Old, What's New, and What's Yet to Come." *Journal of International Economics* 66: 363–84. Summarizes the results we have so far with different theoretical models.

Canzoneri, Matthew, and JoAnna Gray. 1985. "Monetary Policy Games and the Consequences of Noncooperative Behavior." *International Economic Review* 26: 547–64. An early demonstration of the disadvantages of uncoordinated policies.

Cazoneri, Matthew, and Patrick Minford. 1986. "When Coordination Matters: An Empirical Analysis." *Applied Economics* 20: 1137–54. The gains from coordination depend on the size, incidence, and duration of shocks.

Currie, David, Paul Levine, and Nic Vidalis. 1987. "Co operative and Noncooperative Rules for Monetary and Fiscal Policy in an Empirical Two Bloc Model." In *Global Macroeconomics: Policy Conflict and Cooperation*, edited by Ralph Bryant and Richard Portes. London: Macmillan, 73 122. The importance of commitment.

Feldstein, Martin. 1988. "Thinking about International Economic Cooperation." *Journal of Economic Perspectives* 2: 3 13. How coordination could make matters worse.

Foreman Peck, James, Andrew Hughes Hallett, and Yue Ma. 2007. "Trade Wars and the Great Slump." *European Review of Economic History* 11: 73 98. The failure of internal versus international coordination as a cause of the Great Depression of the 1930s.

Hamada, Koichi. 1976. "A Strategic Analysis of Monetary Interdependence." *Journal of Political Economy* 83: 677 700. This paper introduces the importance of international monetary coordination from a theoretical perspective.

Holtham, Gerald, and Andrew Hughes Hallett. 1987. "International Policy Cooperation and Model Uncertainty." In *Global Macroeconomics: Policy Conflict and Cooperation*, edited by Ralph Bryant and Richard Portes. London: Macmillan, 128 77. Uncertainty about the correct model may destroy the gains from coordination.

. 1992. "International Policy Coordination When Policy Makers Do Not Agree on the True Model." *American Economic Review* 82: 1043 51. The form of policy bargain matters when there is uncertainty about how the economy works or about the transmission of policy effects.

Hughes Hallett, Andrew. 1986. "International Policy Design and the Sustainability of Policy Bargains." *Journal of Economic Dynamics and Control* 10: 467 94. The sensitivity of the gains from coordination to different types of policy bargains and cheating.

. 1992. "Target Zones and International Policy Coordination: The Contrast between the Necessary and Sufficient Conditions for Success." *European Economic Review* 36: 893 914. Exchange rate stabilization as a partial substitute for formal policy coordination.

Hughes Hallett, Andrew, Gerald Holtham, and Gary Hutson. 1989. "Exchange Rate Targeting as a Surrogate for International Policy Coordination." In *Blueprints for Exchange Rate Management*, edited by Barry Eichengreen, Marcus Miller, and Richard Portes. New York: Academic Press, 239 78. The distribution of the gains, especially among simplified forms of coordination.

Hughes Hallett, Andrew, and Diana Weymark. 2007. "Government Leadership and Central Bank Design." *Canadian Journal of Economics* 40: 607 27. Fiscal monetary interactions can be used to promote greater internal coordination under the appropriate institutional structures.

Oudiz, Gilles, and Jeffrey Sachs. 1984. "Macroeconomic Coordination among Industrial Economies." *Brookings Papers on Economic Activity* 1: 1 64. The original calculation of the gains from coordination.

Pappa, Eva. 2004. "Do the ECB and the Fed Really Need to Cooperate? Optimal Monetary Policy in a Two Country World." *Journal of Monetary Economics* 51: 753 79. A modern example of monetary coordination.

Rogoff, Kenneth. 1985. "Can International Monetary Policy Coordination Be Counterproductive?" *Journal of International Economics* 18: 199 217. Coordination among a subset or coalition of policymakers can be counterproductive for the whole.

ANDREW HUGHES HALLETT

■ international reserves

International reserves are the liquid external assets (foreign currency, foreign currency bonds, and gold) under the control of the central bank. Under the Bretton Woods system, adequate reserves were measured by months of imports: the prevailing rule of thumb considered four months of imports to be reasonable coverage. This perspective fitted well in a world with limited financial integration, in which trade openness reflected a country's vulnerability to external shocks (Fischer 2001). In the absence of reserves, balance of payment deficits would have to be corrected through a reduction in aggregate expenditures, imposing macroeconomic adjustment costs, manifested in sharp contractions of investment and consumption, thereby inducing recessionary pressures. As greater trade openness increased the exposure to trade shocks, minimizing adjustment costs

required higher reserve holdings. An intriguing development since the 1960s has been that, despite the proliferation of greater exchange rate flexibility, international reserves/gross domestic product (GDP) ratios have increased substantially. Reserve holdings have trended upward; at the end of 1999, reserves were about 6 percent of global GDP, 3.5 times what they were at the end of 1960 and 50 percent higher than in 1990. Practically all the increase in reserves/GDP holding has been by developing countries, mostly concentrated in East Asia (Flood and Marion 2002).

International Reserves as a Buffer Stock The earlier literature focused on using international reserves as a buffer stock, part of the management of an adjustable-peg or managed-floating exchange-rate regime. Accordingly, optimal reserves balance the macroeconomic adjustment costs incurred in the absence of reserves with the opportunity cost of holding reserves (Frenkel and Jovanovic 1981). The buffer stock model predicts that average reserves depend negatively on adjustment costs, the opportunity cost of reserves, and exchange rate flexibility and positively on GDP and reserve volatility, driven frequently by the underlying volatility of international trade. Overall, the literature of the 1980s supported these predictions (see Flood and Marion 2002).

Post-1998 trends in hoarding reserves, especially the large increase in hoarding international reserves in East Asia, stirred lively debate among economists and financial observers. Although useful, the buffer stock model has a limited capacity to account for the recent development in hoarding international reserves the greater flexibility of the exchange rates exhibited after 1990 should work in the direction of reducing reserve hoarding, in contrast to the trends reported earlier. As an indication of excess hoarding, some observers noted that developing countries frequently borrow at much higher interest rates than what they earn on reserves.

International Reserves and Self-Insurance The recent literature provided several interpretations for these puzzles, focusing on the observation that the deeper financial integration of developing countries has increased exposure to volatile short-term inflows of capital (dubbed "hot money"), subject to frequent sudden stops and reversals (see Calvo 1998; Edwards 2004). Looking at the 1980s and 1990s, the magnitude and speed of the reversal of capital flows throughout the 1997 98 East Asian financial crisis surprised most observers (Aizenman and Marion 2003). Most viewed East Asian countries as less vulnerable to the perils associated with hot money than Latin American countries. After all, East Asian countries were more open to international trade, had sounder fiscal policies, and showed much stronger growth performance. In retrospect, the 1997 98 crisis exposed hidden vulnerabilities of East Asian countries, forcing the market to update the probability of sudden stops affecting all countries.

These observations suggest that hoarding international reserves can be viewed as a precautionary adjustment, reflecting the desire for self-insurance against exposure to future sudden stops. Self-insurance has several interpretations. The first focuses on precautionary hoarding of international reserves needed to stabilize fiscal expenditure in developing countries (see Aizenman and Marion 2004). Specifically, a country characterized by volatile output, inelastic demand for fiscal outlays, high tax collection costs, and sovereign risk may want to accumulate both international reserves and external debt (sovereign risk is the added risk, such as default of sovereign governments on debts, and nationalization, assumed by investors with funds invested in foreign counties). External debt allows the country to smooth consumption when output is volatile. International reserves that are beyond the reach of creditors would allow such a country to smooth consumption in the event that adverse shocks trigger a default on foreign debt.

Another version of self-insurance and precautionary demand for international reserves views international reserves as output stabilizers (Ben-Bassat and Gottlieb 1992; Aizenman and Lee 2007). Accordingly, international reserves can reduce the probability of an output drop induced by a sudden stop and/or the depth of the output collapse when the

sudden stop materializes. This argument is in line with the Guidotti-Greenspan rule of thumb of the 1990s — countries should hold liquid reserves equal to their foreign liabilities coming due within a year. This rule reflects the shifting focus from reserve adequacy measured in terms of trade flows of goods to flows of assets.

Back of the envelope estimation suggests that the expected benefits of following a Guidotti-Greenspan rule is about 1 percent of gross domestic product (GDP). This would be the case if a country holding reserves equal to its short-term debt reduces the annual probability of experiencing a sharp reversal in capital flows by 10 percent on average (in line with Rodrik and Velasco 1999, see Rodrik 2006) and if the output cost of a financial crisis is about 10 percent of GDP, as found by Hutchison and Noy (2006). Similar results have been obtained using more elaborated models (see Garcia and Soto 2004; Jeanne and Ranciere 2005). These authors concluded that self-insurance against sudden stops plays an important role in accounting for recent hoarding of international reserves.

While the Guidotti-Greenspan-IMF rule focused on the ratio of reserves to short-term debt, Kim et al. (2005) looked at a more flexible rule, based on the behavior of different types of capital flows during currency crises. Application to selected Asian countries leads them to conclude that the countries affected by the East Asian crisis held excessive reserves by 2003 — the affected countries have already built up more than adequate reserve levels to handle a repeat of the actual capital outflows that occurred during the 1997–98 crises scaled up to 2003 values. One may note, however, that the rapidly changing structure of the developing countries' financial integration implies that future possible crises would not resemble the previous ones. For example, Korea, one of the countries affected by the 1997–98 crisis, lifted restrictions on foreign equity ownership in the aftermath of the crisis. In response, foreigners' shareholding as a percentage of the total market capitalization has risen from 12 percent in 1997 to 40 percent by 2003. Arguably, the sizable accumulation of reserves by Korea during that period may reflect

the wish to cover short-term external debt plus some portion of foreigners' shareholdings, in the desire to reduce possible real exchange rate repercussion of future reversals of capital flows.

International Reserves: Precaution versus Mercantilism The Korean policy suggests another angle associated with international reserves — the possibility that international reserves management may lower real exchange rate volatility, which in turn may allow a smoother output and potentially higher growth rate. To put this topic in broader context, note that the literature of the 1990s identified large adverse effects of exogenous volatility on the GDP and economic growth in developing countries. An important channel that may explain such negative levels and growth effects of volatility are capital market imperfection and low levels of financial development (Aghion et al. 2006).

The views linking the large increase in hoarding reserves to growing exposure to sudden stops associated with financial integration face a well-known contender in a modern incarnation of mercantilism (Dooley et al. 2003). According to this interpretation, reserves accumulation is a by-product of promoting exports, which is needed to create better jobs, thereby absorbing abundant labor in traditional sectors. Though intellectually intriguing, this interpretation remains debatable — the history of Japan and Korea suggests the near absence of mercantilist hoarding of international reserves during the phase of fast growth, and the prevalence of export promotion by preferential financing in targeted sectors. Foundering economic growth led to the onset of large hoarding of reserves both in Japan and Korea, probably due to both mercantilist motives and self-insurance to deal with growing fragility of the banking system. These perspectives suggest that the massive hoarding of reserves by China is a hybrid of the mercantilist and self-insurance motives (Aizenman and Lee 2006). Yet mercantilist hoarding by one country may induce competitive hoarding by other countries to preempt any competitive advantage gained by the first country, a reaction that would dissipate most competitiveness gains. This view is supported by the interdependence of the demand for

international reserves among ten East Asian countries (Cheung and Qian 2006).

Overall, greater exposures of developing countries to sudden stops and reversals of hot money, growing trade openness, and the desire to improve competitiveness and to reduce real exchange rate volatility go a long way toward accounting for the observed increase in the rapid and massive stockpiling of international reserves by developing markets.

See also Bretton Woods system; currency crisis; dollar standard; dominant currency; exchange rate regimes; exchange rate volatility; financial crisis; foreign exchange intervention; global imbalances; gold standard, international; hot money and sudden stops; international liquidity; mercantilism; real exchange rate; reserve currency; sterilization; vehicle currency

FURTHER READING

Aghion, Philippe, Philippe Bacchetta, Romain Ranciere, and Kenneth Rogoff. 2006. "Exchange Rate Volatility and Productivity Growth: The Role of Financial Development." NBER Working paper No. 12117. Cambridge, MA: National Bureau of Economic Research. Evidence that real exchange rate volatility is associated with lower productivity growth.

Aizenman, Joshua, and Jaewoo Lee. 2006 "Financial versus Monetary Mercantilism Long run View of Large International Reserves Hoarding." NBER Working Paper No. 12718. Cambridge, MA: National Bureau of Economic Research. A key instrument of the East Asian development strategy has been subsidizing the cost of capital; hoarding international reserves tend to occur in the aftermath of floundering growth; defines and explains competitive hoarding.

. 2007. "International Reserves: Precautionary versus Mercantilist Views, Theory and Evidence." *Open Economies Review* 18 (2): 191 214. Evidence and a model of the relative importance of the precautionary and mercantilist views of international reserves.

Aizenman, Joshua, and Nancy P. Marion. 2003. "The High Demand for International Reserves in the Far East: What's Going On?" *Journal of the Japanese and International Economies* 17 (3): 370 400. Explains the higher demand for reserves using the precautionary demand and downside risk aversion motives.

. 2004. "International Reserves Holdings with Sovereign Risk and Costly Tax Collection." *Economic Journal* 114 (July): 569 91. Models precautionary use of reserves by developing countries and the negative impact of political instability on international reserves.

Ben Bassat, Avraham, and Daniel Gottlieb. 1992. "Optimal International Reserves and Sovereign Risk." *Journal of International Economics* 33: 345 62. Explains and estimates the importance of sovereign risk and the cost of default as a major determinant of the demand for international reserves.

Calvo, Guillermo. 1998. "Capital Flows and Capital market Crises: The Simple Economics of Sudden Stops." *Journal of Applied Economics* 1 (1): 35 54. Explains mechanisms through which a sudden stop in international credit flows may bring about financial and balance of payments crises and examines factors triggering sudden stops.

Cheung, Yin Wong, and Xing Wang Qian. 2006. "Hoarding of International Reserves: Mrs. Machlup's Wardrobe and the Joneses." Manuscript. Santa Cruz: University of California at Santa Cruz. Shows the interdependence of the demand for international reserves among 10 East Asian countries.

Dooley, P. Michael, David Folkerts Landau, and Peter Garber. 2003. "An Essay on the Revived Bretton Woods System." NBER Working Paper No. 9971. Cambridge, MA: National Bureau of Economic Research. The use of fixed exchange rate in Asia has reestablished the United States as the center country in the modified Bretton Woods system; links it to an export led growth supported by undervalued exchange rates strategy of East Asia.

Edwards, S. 2004. "Thirty Years of Current Account Imbalances, Current Account Reversals, and Sudden Stops." *IMF Staff Papers* 51 (special issue): 1 49. Major reversals in current account deficits have been associated to sudden stops of capital inflows; the probability of a reversal depends on external debt/GDP, international reserves/GDP, and debt services. More open countries suffer less in a reversal.

Fischer, Stanley. 2001. "Opening Remarks." IMF/World Bank International Reserves: Policy Issues Forum. Washington, DC. Overview of the changing patterns of hoarding international reserves.

Flood, Robert, and Nancy P. Marion. 2002. "Holding In ternational Reserves in an Era of High Capital Mobi lity." In *Brookings Trade Forum 2001*, edited by S. Collins and D. Rodrik. Washington, DC: Brookings Institution Press. Buffer stock reserve models work about as well in the modern floating rate period as they did during the Bretton Woods regime. During both periods models' fundamentals explain only about 15 percent of reserves volatility.

Frenkel, Jacob, and Boyan Jovanovic. 1981. "Optimal International Reserves: A Stochastic Framework." *Economic Journal* 91 (362): 507 14. Adjustment cost model of the demand for international reserves, high lighting the role of uncertainty.

Garcia, Pablo, and Claudio Soto. 2004. "Large Holdings of International Reserves: Are They Worth It?" Central Bank of Chile Working Papers No. 299 (December). Santiago: CBC. Estimate crisis probabilities and apply it to evaluate the optimal stock of reserves, concluding that observed stocks of reserves for most of the cases are consistent with an optimal self insurance policy.

Hutchison, M. Michael, and Ilan Noy. 2006. "Sudden Stops and the Mexican Wave: Currency Crises, Capital Flow Reversals, and Output Loss in Emerging Markets." *Journal of Development Economics* 79 (1): 225 48. Sudden stop crises have a large negative, but short lived, impact on output growth over and above that found with currency crises. The cumulative output loss of a sudden stop is around 13 15 percent over a 3 year period.

Jeanne, Olivier, and Romain Ranciere. 2005. "The Optimal Level of International Reserves for Emerging Market Economies: Formulas and Applications." Washington, DC: IMF Research Department. An insurance model of international reserves in the presence of sudden stop can explain reserves of the order of magnitude observed in many emerging market countries, but not the recent buildup of reserves in Asia.

Kim, Jung Sik, Jie Li, Ramkishen Rajan, Ozan Sula, and Thomas D. Willett. 2005. "Reserve Adequacy in Asia Revisited: New Benchmarks Based on the Size and Composition of Capital Flows." In *Claremont KIEP Conference on Monetary and Exchange Rate Arrangements in East Asia*. Korean Institute of International Economic Policy, Seoul. Presents new benchmarks for judging re serve adequacy and concludes that the reserve levels in East Asia are more than that needed to finance capital outflows of the severity of the 1997 98 Asian crises.

Rodrik, Dani. 2006. "The Social Cost of Foreign Exchange Reserves." *International Economic Journal* 20 (3): 253 66. Cost benefit analysis of hoarding international re serves using the self insurance motive. Explaining why developing countries have not tried harder to reduce short term foreign liabilities instead of hoarding reserves remains a puzzle.

Rodrik, Dani, and Andres Velasco. 1999. "Short Term Capital Flows." NBER Working Paper No. 7364. Cambridge, MA: National Bureau of Economic Re search. A model of self fulfilling crises; estimation shows that greater short term debt/reserves ratio is a robust predictor of financial crises and is associated with more severe crises when capital flows reverse.

JOSHUA AIZENMAN

■ **international risk sharing**

See capital mobility

■ **intrafirm trade**

Intrafirm trade is international trade that occurs be tween different affiliates of the same multinational firm. In the United States, intrafirm trade accounted for approximately 35 percent of all international trade in 2004, and it has been a large fraction of international trade in the United States for decades. In the 2002 *Economic Outlook* of the Organisation for Economic Co-operation and Development there is some evidence of increases in intrafirm trade in other countries.

Data on intrafirm trade are difficult to find con sistently across countries. The U.S. data come from the Bureau of Economic Analysis, which undertakes annual surveys of U.S. multinational firms and their affiliates abroad as well as foreign-based multina tional firms and their affiliated firms in the United States.

The large magnitude of intrafirm trade has been emphasized by many scholars of foreign direct investment (FDI). Furthermore, many scholars have discussed the increasing importance of vertical specialization in international trade. Vertical specialization includes the segmentation of production across national boundaries, whereby different stages of the production process are done in different countries, in line with their natural comparative advantage. Yi (2003) argues that vertical specialization can help explain the growth in world trade. Still, it is important to distinguish trade in intermediate inputs from intrafirm trade, as the former need not occur within the firm.

Understanding the nature of intrafirm trade therefore requires understanding the motives behind FDI more generally, or the rationale for organizing cross-border economic activity within the firm. Theoretical models of multinational activity emphasize several possible motives for internalizing activity within the firm. For example, there may be a need to control the quality of the product, accompanied by difficulties in the formation of contracts at arm's length to ensure the reputation of the firm. Also, proprietary firm-specific knowledge can make it difficult to appropriate the gains from production via licensing, as it is difficult to charge the appropriate fee for the knowledge without revealing the knowledge itself, thus lowering the incentive to pay for it. There is evidence in favor of such motivations. For example, the literature has consistently shown a correlation between the propensity to exchange goods within the firm and the research and development intensity of production; see Borga and Zeile (2004) for an overview of findings in this area.

An additional motive for internalization is tax avoidance. Gordon and Hines (2002) undertake a comprehensive survey of international taxation and find little evidence for viewing multinational firms as financial intermediaries, facilitating the movement of capital to higher return locations. Instead, they find more support for models of multinational firms as facilitators of tax avoidance or owners of firm-specific intangible capital.

Countries tax multinational firms in a variety of ways. Some countries exempt the foreign earnings of their resident firms from taxation; these are referred to as territorial systems. Multinational firms based in such countries have an unambiguous incentive to shift income to low-tax locations. Other countries (such as the United States) tax the worldwide income of their resident firms, but they still generally allow deferral of taxation on income earned in low-tax countries until that income is repatriated. These are referred to as credit systems. Deferral, as well as other methods of lowering the residual tax due, provides multinational firms in credit system countries with a substantial incentive to shift profits to low-tax countries.

Intrafirm trade may facilitate international tax avoidance by multinational firms in several ways. By altering the prices of transactions that occur within the firm, firms can effectively shift income from high-tax locations to low-tax locations. For example, given these tax incentives, multinational firms have an incentive to overprice intrafirm trade flows originating in low-tax countries and underprice intrafirm trade flows originating in high-tax countries. Evidence that firms follow such strategies is provided in Clausing (2003), which examines intrafirm trade prices directly. Data on intrafirm trade prices are uncommon, but most of the literature in this area provides consistent, if indirect, empirical support for the importance of tax-motivated transfer pricing. Hines (1999) reviews this body of literature.

In addition to tax distortions on the pricing of intrafirm transactions, there is evidence (e.g., Clausing 2006 and references within) that firms undertake greater quantities of intrafirm trade with low-tax countries. This finding may indicate that low-tax locations are attractive places to undertake FDI, and thus intrafirm trade is also encouraged as it is complementary to multinational activity. This finding may also indicate that greater quantities of intrafirm trade are undertaken in order to facilitate tax avoidance; these two motives are difficult to distinguish given data constraints.

Aside from the tax influences on intrafirm trade, the behavior of intrafirm trade more generally has also received attention in the literature. Some scholars have argued that intrafirm trade by multinational enterprises may be more responsive to changes in economic conditions than international trade between unaffiliated entities. Feinberg and Keane (2006, 2007) consider the relationship between changes in technology, tariff reductions, the restructuring of multinational firm production, and the growth of world trade and intrafirm trade. They find that the growth in Canadian-U.S. intrafirm trade is due in large part to technical change and in particular improvements in logistics management and "just-in-time" production systems. Tariff reductions play a relatively minor role in intrafirm trade, although they are important in explaining arm's-length trade by multinational firms. So the reasons for growth in intrafirm trade appear to differ from the reasons for the growth in arm's-length trade by multinational firms.

International trade by multinational firms is a majority of all trade for the United States, and it is likely to be of similar importance for many other countries. While some of this trade occurs at arm's length among unaffiliated firms, an important part of the international trade of multinational firms is intrafirm trade. Intrafirm trade, or trade between different affiliates of the same multinational firm, may differ from conventional trade in important ways. Foremost, it is influenced by international tax incentives, but it also may respond differently to other key economic variables.

See also foreign direct investment (FDI); multinational enterprises

FURTHER READING

Borga, Maria, and William J. Zeile. 2004. "International Fragmentation of Production and the International Trade of U.S. Multinational Companies." Bureau of Economic Analysis Working Paper No. 2004 02. Washington, DC: BEA. Discussion of intrafirm trade.

Clausing, Kimberly A. 2006. "International Tax Avoidance and U.S. International Trade." *National Tax Journal* 59 (2): 269 87. Provides evidence regarding tax incentives and intrafirm trade.

———. 2003. "Tax Motivated Transfer Pricing and U.S. Intrafirm Trade Prices." *Journal of Public Economics* 87 (9 10): 2207 23. Provides evidence regarding tax motivated transfer pricing and intrafirm trade prices.

Feinberg, Susan E., and Michael P. Keane. 2006. "Accounting for the Growth of MNC Based Trade Using a Structural Model of U.S. MNCs." *American Economic Review* 96 (5): 1515 58. Evidence that growth in U.S. Canadian intrafirm trade was due to technical change, not tariff reductions.

———. 2007. "Advances in Logistics and the Growth of Intrafirm Trade: The Case of Canadian Affiliates of U.S. Multinationals, 1984 1995." *Journal of Industrial Economics* 55 (4): 571 632. Shows that improved logistics, including "just in time" production, explains much of growth in intrafirm trade.

Gordon, Roger H., and James R. Hines Jr. 2002. "International Taxation." In *Handbook of Public Economics*, vol. 4, edited by Alan Auerbach and Martin Feldstein. North Holland, 1935 95. Overview of international taxation.

Hines, James R. Jr. 1999. "Lessons from Behavioral Responses to International Taxation." *National Tax Journal* 52 (2): 305 22. Overview of international taxation literature, with an emphasis on empirical work.

Yi, Kei Mu. 2003. "Can Vertical Specialization Explain the Growth of World Trade?" *Journal of Political Economy* 111 (1): 52 102. Examines the contribution of vertical specialization to the growth of world trade.

KIMBERLY A. CLAUSING

■ **intraindustry trade**

Intraindustry trade arises if a country simultaneously imports *and* exports similar types of goods or services. Similarity is identified here by the goods or services being classified in the same sector. Suppose, for the sake of argument, that we focus on the sector "cars." Intraindustry trade then occurs, for example, if Germany exports cars to France and simultaneously imports cars from Italy. On the one hand this raises the question why Germany is (at least partially) ex-

porting cars in exchange for importing cars instead of focusing exclusively on so-called *inter*industry trade, namely exporting cars in exchange for importing different types of goods (such as food or airplanes). On the other hand, this raises the question why different goods are lumped together in the same sector, as the exported Volkswagen Golfs differ from the imported Ferraris. We address these two basic questions in what follows.

Although in hindsight various antecedents can be traced, the phenomenon of intraindustry trade as such first received attention in the 1960s in studies by Pieter Verdoorn and Bela Balassa on the increased trade flows among European countries. Herbert Grubel and Peter Lloyd (1975) provided the definitive empirical study on the importance of intraindustry trade and how to measure it. Solid theoretical foundations for explaining intraindustry trade came later (in the 1980s and 1990s) with the new trade literature, to a large extent based on a monopolistic competition framework.

Types of Intraindustry Trade It is customary to distinguish between two different types of intraindustry trade, each warranting a different type of explanation, namely:

Horizontal intraindustry trade. This refers to the simultaneous exports and imports of goods classified in the same sector and at the *same stage of processing*. This is likely based on product differentiation, for example, South Korea's simultaneous import and export of mobile telephones in the final processing stage. As these mobile phones are produced using similar technology and provide similar functions they are classified in the same sector. Nonetheless, the exported Samsung telephones differ in appearance and product characteristics slightly from the imported Nokia telephones, catering to the desires of different types of consumers.

Vertical intraindustry trade. This refers to the simultaneous exports and imports of goods classified in the same sector but at *different stages of processing*. This is likely based on the increasing ability to organize fragmentation of the production process into different stages, each performed at different locations by taking advantage of the local conditions. China,

for example, imports technology-intensive computer components and uses its abundant, low-wage labor force to assemble these components in the labor-intensive final production stage, before the components (as part of a finished computer) are exported again to Europe or the United States.

Measuring Intraindustry Trade: The Grubel-Lloyd Index The most often used method for determining the extent of intraindustry trade was proposed by Grubel and Lloyd (1975). This measure, now known as the Grubel-Lloyd index, is simple to calculate and intuitively appealing. Once a country's export and import value for a particular sector and period are known, it is calculated as:

$$GL_{sector\ i} = 1 - \left(\frac{|export_{sector\ i} - import_{sector\ i}|}{export_{sector\ i} + import_{sector\ i}} \right). \tag{1}$$

If the country only imports or only exports goods or services within the same sector, such that there is no intraindustry trade, the second term on the right-hand side of equation (1) is equal to one, such that the whole expression reduces to zero. Similarly, if the export value is exactly equal to the import value ($export_{sector\ i} = import_{sector\ i}$), the second term on the right-hand side of equation (1) is equal to zero, such that the whole expression reduces to one. The Grubel-Lloyd index therefore varies between zero (indicating pure *inter*industry trade) and one (indicating pure *intra*industry trade).

Data Aggregation As already indicated, to some extent intraindustry trade can be considered a classification problem as different types of goods and services are lumped together in the same sector. In practice, international trade flows are classified in various ways. Using the SITC (standard international trade classification) we can distinguish ten different broad sectors (the so-called one-digit level). Each of these one-digit sectors can, in principle, be subdivided into ten more detailed two-digit sectors. Each of the two-digit sectors can in turn, in principle, be subdivided into ten even more detailed three-digit sectors, and so forth. Sector 6 at the one-digit level, for example, consists of "manufactured goods." One of the subsectors at the two-digit level is sector 61

"leather manufactures" while another is sector 63 "cork/wood manufactures." Analyzing intraindustry trade at the very broad one-digit level therefore classifies trade of leather manufactures in exchange for cork/wood manufactures as intraindustry trade, which seems unwarranted. Looking at the more detailed two-digit level this problem partially disappears and a smaller extent of trade is therefore classified as intraindustry trade. A further reduction occurs if we look at even more detailed levels of aggregation. The three-digit level, for example, distinguishes between cork manufacturers (sector 633) and different types of wood manufacturers (sectors 634 and 635) separately.

Table 1 illustrates the data aggregation problem for China for three different levels of aggregation (consisting of 10, 67, and 237 different sectors) by reporting a trade-weighted average Grubel-Lloyd index for a selection of years. We report the one-digit level only for the sake of argument as it is generally considered too crude a classification in practice. Three things are clear from the table. First, as we distinguish between more sectors a smaller fraction of trade is classified as intraindustry trade, for example reducing it from 58 percent to 49 percent to 42 percent of total trade in China in 2005. Second, even though intraindustry trade reduces as we identify more sectors, it does not disappear. It is, for example,

still 42 percent of total trade at the three-digit level in China in 2005. This is a general characteristic of current trade flows as intraindustry trade exists for very detailed sector classifications. Third, and most important, intraindustry trade seems to become more important over time, for example increasing at the three-digit level in China from 20 percent in 1980 to 42 percent in 2005. We now turn to this issue.

Empirical Characteristics of Intraindustry Trade
There are structural differences across sectors regarding the extent of intraindustry trade. To demonstrate this, we use the factor-intensity classification of the International Trade Center, a joint agency of the United Nations Conference on Trade and Development and the World Trade Organization, which distinguishes between five broad factor-intensity categories at the three-digit level, namely (within parentheses the number of the sector belonging to the particular category):

Primary products (83): e.g. meat, dairy, cereals, fruit, coffee, minerals, and oil.
Natural-resource-intensive products (21): e.g. leather, wood, pig iron, and copper.
Unskilled-labor-intensive products (26): e.g. textiles, clothing, ships, and footwear.
Human-capital-intensive products (43): e.g. perfumes, cosmetics, cars, and watches.
Technology-intensive products (62): e.g. chemicals, electronics, tools, and aircraft.

Table 2 depicts the extent of intraindustry trade for these different types of sectors in China for selected years. It shows that the level of intraindustry trade is particularly low for unskilled-labor-intensive sectors, particularly high for technology-intensive sectors, and intermediate for the other types of sectors. As countries such as China successfully develop, the composition of their trade flows tends to move away from primary products, initially toward unskilled-labor-intensive products and subsequently toward technology- and human-capital-intensive products. Associated with these changes there is an ultimate increase in the extent of intraindustry trade.

Table 1
Intraindustry trade and aggregation: China, selected years (trade weighted average Grubel Lloyd index, different levels of aggregation)

	3 digit 237 sectors	2 digit 67 sectors	1 digit 10 sectors
1980	0.20	0.30	0.63
1985	0.20	0.29	0.44
1990	0.36	0.45	0.60
1995	0.38	0.48	0.67
2000	0.39	0.48	0.57
2005	0.42	0.49	0.58

Source: Author's calculations based on United Nations (2006) COMTRADE data from World Integrated Trade Solution (WITS), Geneva.

Table 2
Intraindustry trade and composition of trade flows: China, selected years
(trade weighted average Grubel Lloyd index [3 digit level] and percent of total trade)

	Type of products				
	Primary products	Natural resource intensive	Unskilled labor intensive	Technology intensive	Human capital intensive
Weighted average Grubel Lloyd summary statistics for product type, 1980 2005					
average	0.27	0.38	0.16	0.56	0.36
st dev[a]	0.11	0.07	0.04	0.04	0.08
Share of product type in total trade (percent)					
1980	51.4	3.4	27.8	8.1	9.2
1985	49.5	2.0	33.7	7.1	7.7
1990	19.4	2.9	46.5	15.6	15.5
1995	10.1	4.0	45.4	24.9	15.6
2000	7.5	3.2	39.2	35.5	14.6
2005	4.6	3.3	28.9	47.7	15.5

Source: Author's calculations based on United Nations (2006) COMTRADE data from World Integrated Trade Solution (WITS), Geneva.
[a] st dev = standard deviation.

In a study summarizing the growing importance of intraindustry trade, the Organisation for Economic Co-operation and Development (OECD 2002) lists the following empirical characteristics: Intraindustry trade

- has risen significantly since the 1980s in most (OECD) countries
- is particularly high for sophisticated manufactured products (chemicals, machinery, transportation equipment, electrical equipment, and electronics; both based on product differentiation and fragmentation)
- is particularly high for very open countries ("supertrading" economies, where both imports and exports account for more than half of GDP)
- is connected to foreign direct investment inflows, particularly in Eastern European "transition" economies
- is related to preferential trade agreements, for example, the sharp increase in in-

traindustry trade in Mexico after the North American Free Trade Agreement
- is to a large extent based on intrafirm trade, either based on product variety or on fragmentation (intrafirm trade accounts, for example, for one-third of exports in Japan and the United States).

See also economies of scale; fragmentation; Heckscher-Ohlin model; intrafirm trade; monopolistic competition; New Trade Theory; vertical versus horizontal foreign direct investment

FURTHER READING

Grubel, Herbert G., and Peter Lloyd. 1975. *Intraindustry Trade: The Theory and Measurement of International Trade in Differentiated Products*. London: Macmillan. Most influential study on measuring and understanding intraindustry trade.

Helpman, Elhanan, and Paul Krugman. 1987. *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*. Boston:

MIT Press. Overview of the main theories explaining intraindustry trade by two prominent contributors.

Organisation for Economic Co operation and Develop ment (OECD). 2002. "Intraindustry and Intrafirm Trade and the Internationalisation of Production." *Economic Outlook*, no. 71, chap. 6: 159 70. A recent discussion of the growing importance and characteristics of intraindustry trade.

United Nations (UN). 2006. COMTRADE data from World Integrated Trade Solution (WITS). Geneva. Available at https://unp.un.org/comtrade.aspx. A large trade data set from over 200 reporting countries or areas, covering 45 years of data and more than 5,000 different products.

CHARLES VAN MARREWIJK

■ J-curve effect

The J-curve effect describes the time lag with which a currency depreciation or devaluation leads to an improvement in the trade balance. Although the trade balance may improve in the long run, it may worsen initially so that it follows the pattern of a *J* tilted to the right. The origin of the J-curve is usually attributed to economist Stephen Magee's (1973) attempt to find an explanation for the short-run behavior of the U.S. trade balance in the early 1970s. In 1971 a balance of trade surplus turned into a deficit, and although the U.S. dollar was devalued, the trade balance continued to deteriorate.

The theoretical basis of the J-curve effect is the elasticities approach to the balance of payments. According to this theory a currency devaluation or depreciation is expected to improve the trade balance by changing the relative prices of domestic and foreign goods. By making foreign goods more expensive in the home country and the home country's goods cheaper abroad, demand for imports will be low and foreigners will buy more of the home country's exports. Provided that the responses of importers and exporters to the price changes are strong enough, the Marshall-Lerner condition will be fulfilled and the trade balance will improve.

Importance of the J-Curve The J-curve effect is important because it has implications for the effectiveness of policies designed to improve the balance of payments. During the 1997 financial crisis in Asia, for example, the International Monetary Fund encouraged currency depreciation as part of a package of policies to stabilize the balance of payments of countries hardest hit by the crisis. If the J-curve effect is important then it will take a longer period of time before the balance of payments improves than if the adjustment to the currency depreciation had been instantaneous.

The J-curve phenomenon has also been invoked to explain the persistence of the U.S. trade deficit following the fall in the U.S. dollar from its peak in 1985. Even after two years of a falling dollar, the trade deficit failed to show a marked improvement. According to this perspective, while the U.S. trade balance was expected to improve in the longer term, the substantial lags in the adjustment of both prices and quantities to exchange rate changes contributed to the sluggish response of the U.S. deficit to the U.S. dollar depreciation.

Causes of the J-Curve The causes of the J-curve effect are generally traced to a currency contract effect and/or weak short-run responses to changes in the relative prices of exports and imports following a depreciation or devaluation of the home country's currency. The currency contract effect arises when import and export orders reflect decisions made in advance of the devaluation at the old exchange rate. Therefore, immediately after a devaluation there may be a period when contracts signed prior to the exchange rate change become due. The J-curve can then occur if the home country's exports are invoiced in the home currency but its imports are invoiced in the foreign currency: during the currency contract period the value of exports in the home country's currency is unaffected by the devaluation, but import values measured in home currency will rise, thus

worsening the trade balance. Only in the longer run when new contracts are signed will the value of home imports fall and home exports rise sufficiently to improve the trade balance.

Even in the absence of a currency contract period, a J-curve can occur if responses to the changes in import and export prices are weak, so that quantities traded do not change much in the short run. This would imply a failure of the Marshall-Lerner condition over this period. If the demand for imports is unresponsive, the total import bill could rise until the home country's buyers find suitable substitutes, while the demand for exports may be insensitive if foreign buyers do not increase their demand for the home country's goods in the short run even though their prices have fallen in the foreign country's currency.

Estimating the J-Curve Although economists have done a substantial amount of empirical work on the J-curve, the results seem to depend on the particular country and the period selected for investigation. What is clear is that the short-run response of the trade balance to a devaluation or depreciation of the currency does not follow a uniform pattern. One reason for this is that other factors may be involved. For example, the rapid improvement in the Mexican trade balance with the United States following the sharp depreciation of the peso against the dollar in December 1994 is often cited as a successful balance of payments adjustment with the absence of any obvious J-curve complications. But the massive size of the devaluation and the financial crisis that accompanied it may have persuaded exporters and importers that the devaluation was permanent and therefore speeded up their adjustment. In addition, an improvement in the trade balance may have resulted from a fall in spending on imports as government spending was cut, and prior trade liberalization may have made import and export volumes more sensitive to relative price changes.

A second factor is that the response to the exchange rate change may vary according to the level of sectoral aggregation. Countries with a high proportion of natural resources, such as crude petroleum and agricultural products, in their exports and/or imports may find their trade balance is less sensitive to exchange rate changes than countries that predominantly trade manufactured goods. A further problem is the need to look beneath the aggregate trade balance to see what is happening at the bilateral level. The trade balance may be improving in the short run with respect to one trading partner but simultaneously deteriorating with respect to another.

See also balance of payments; currency crisis; effective exchange rate; exchange rate pass-through; financial crisis; Marshall-Lerner condition; real exchange rate; vehicle currency

FURTHER READING

Bahmani Oskooee, Mohsen, and A. Ratha. 2004. "The J Curve: A Literature Review." *Applied Economics* 36: 1377 98. Provides a recent review of the vast empirical literature on the J curve effect.

Krugman, P. 1989. "The J Curve, the Fire Sale, and the Hard Landing." *American Economic Review* 79: 31 35. An interesting attempt to unravel the reasons for the sluggish improvement in the U.S. trade balance following the sharp fall in the U.S. dollar in 1985.

Magee, Stephen. 1973. "Currency Contracts, Pass Through, and Devaluation." *Brookings Papers on Economic Activity* 1: 303 25. The pioneering work on the J curve effect.

PETER WILSON

■ joint ventures

A joint venture is a mechanism for combining complementary assets owned by separate firms. These assets can be tangible, such as machinery and equipment, or intangible, such as technological know-how, production or marketing skills, brand names, and market-specific information. In an equity joint venture the partner firms transfer all or part of their assets to a legally independent entity and share the profits from the venture. Contractual arrangements that do not involve shared equity control are sometimes referred to as nonequity joint ventures; examples include licensing and management contracts, as well as supply and distribution agree-

ments. Shared ownership and contractual arrangements are also frequently grouped together under the term *alliances*. This entry focuses on equity joint ventures, specifically international joint ventures involving partners from different countries.

From a world economy perspective there are at least two reasons for examining international joint ventures. First, international joint ventures represent a form of foreign direct investment (FDI). Multinational enterprises often have to decide whether to wholly own a foreign affiliate or to share equity control with a local partner. This decision is a key element of the foreign investment strategy. Second, the ownership structure of a foreign investment project affects host-country welfare. A direct effect comes from the sharing of profits between the multinational and the local firm. Indirect effects arise because ownership influences investors' incentives to commit resources to the project, such as capital and technology. Some host countries impose local ownership requirements that limit the equity stake foreign investors can take in local companies. This raises the question of what the economic effects of such requirements are.

Probably the most comprehensive data on international joint ventures come from the U.S. Department of Commerce benchmark surveys. Although they are not representative for the world as a whole, these data still offer information on long-run trends in international joint venture activity for one of the most important source countries of FDI. According to Desai, Foley, and Hines (2004), who have examined these data, about 80 percent of all U.S. affiliates abroad in 1997 were wholly owned, with the remaining 20 percent equally divided between minority- and majority-owned affiliates. The ownership share is positively correlated with host-country gross national product (GNP). In the richest quartile of host countries, partially owned affiliates accounted for only 15.5 percent, whereas in the poorest quartile they made up more than half of all U.S. foreign affiliates. The data exhibit considerable variation in ownership shares across industries and over time. Since the mid-1980s there has been a downward trend in minority-owned and an upward trend in

majority- and wholly owned affiliates, partly due to changes in U.S. tax laws.

Profit-Maximizing Joint Ventures Consider a multinational firm contemplating a foreign investment project that requires a combination of its own assets and those of a local firm in the host country. What is the appropriate ownership structure of the project? The extensive literature on this issue starts from the premise that the ownership structure is a response to the presence of market failures (or high transaction costs) in asset markets. This can best be understood by assuming counterfactually that there are no such failures. This is the case, if (1) all assets and other inputs to and outputs from the project are observable and verifiable by third parties, such as courts; (2) it is possible to write contracts specifying the provision of each input, and the distribution of output and profits under all possible contingencies; and (3) these contracts can be enforced at no cost. Under these ideal conditions, the ownership structure is indeterminate, since the firms can simply use contracts to coordinate the use of their assets.

Such ideal conditions are unlikely to prevail in practice. Suppose, for example, that the project requires a combination of the multinational's production technology and the local firm's marketing know-how. It may be very difficult to specify what these assets entail, to assess how valuable each asset will be for the project, and hence to write a contract on what each party has to contribute and how profits are to be shared. Even if the two parties both knew how important the technology and the marketing know-how were, it would be next to impossible for a third party to verify this and hence to determine whether both parties have fulfilled their contractual obligations. It may also be impossible through contractual means to prevent spillovers of the technology to the local firm, which could then use it for its own purposes. Specifying appropriate management contracts to ensure that profits from the project are maximized may be difficult, especially if monitoring costs are high. In short, contracts will generally be insufficient to prevent opportunistic behavior. Retaining (partial) ownership of assets, and hence

residual rights of control over them, may then be preferable because it ensures that a firm will obtain at least some return from the project and hence have an incentive to contribute assets to the project and provide effort in running it.

Shared ownership of an investment project is only a second-best solution. First, it may be impossible to guarantee each party the full return from the use of its assets. Hence there may be too little provision of assets or insufficient investment in tailoring the assets to the project. Second, joint ventures require significant management resources due to the need to coordinate decisions between the partners. Why then would one of the partners, say the multinational, not simply acquire the other and assume whole ownership of the project? One obvious advantage of shared ownership is that it requires less capital than a complete takeover. Moreover, the multinational may be interested in only some of the assets of the local firm. If these are hard to disentangle from the local firm's other assets, a joint venture may be the better option. Incomplete information about the value of the local firm's assets provides another reason for shared ownership, since letting the local firm choose how much ownership to retain may reveal information to the multinational.

Empirically it is difficult to distinguish between different explanations for shared ownership, since the information on which firms base their decisions is often confidential. In addition, the ownership decision may be made simultaneously with other decisions concerning the firm's operations. Desai, Foley, and Hines (2004) therefore use changes in U.S. tax laws affecting ownership and the liberalization of local ownership requirements to identify possible interactions between these decisions. They find a complementary relationship between the ownership share of the multinational and the amount of intrafirm trade between the affiliate and the parent company. Firms that trade more internally are more likely to have whole or majority ownership, whereas affiliates selling more of their output or buying more of their inputs locally are more likely to be organized as joint ventures. Possible reasons for this are that

whole ownership reduces the cost of coordinating intrafirm transactions and makes it easier to set internal transfer prices to avoid taxes.

Host-Country Policy Local ownership requirements are most frequently imposed by developing and transition countries, although some high-income countries also put limits on foreign ownership in certain sectors. Possible economic rationales for requiring local equity participation in low-income countries are that they might facilitate spillovers of technology and management know-how to local firms and might secure a share of the project's earnings for the host country when the fiscal system is too inefficient to do this directly through taxes. Some authors argue that multinationals, too, may have an interest to take on a local partner to smooth relations with the host-country government and reduce the risk of expropriation.

See also foreign direct investment (FDI); mergers and acquisitions

FURTHER READING

Asiedu, Elizabeth, and Hadi S. Esfahani. 2001. "Ownership Structure in Foreign Direct Investment Projects." *Review of Economics and Statistics* 83 (4): 647 62. Shows that foreign equity shares rise with the importance of foreign investor assets and decline with the contribution of local assets toward the surplus generated by the project.

Caves, Richard E. 2007. *Multinational Enterprise and Economic Analysis*. 3d ed. Cambridge: Cambridge University Press. A classic examination of the economics of foreign direct investment.

Desai, Mihir A., C. Fritz Foley, and James R. Hines. 2004. "The Costs of Shared Ownership: Evidence from International Joint Ventures." *Journal of Financial Economics* 73 (2): 323 74. Finds that whole ownership is most common when firms coordinate production activities across locations, transfer technology, and implement global tax planning. Observes that much of decline in use of joint ventures can be attributed to increased importance on intrafirm transactions.

Müller, Thomas, and Monika Schnitzer. 2006. "Technology Transfer and Spillovers in International Joint Ventures." *Journal of International Economics* 68 (2):

456 68. Shows why, on account of host country poli
cies, a multinational might enter a joint venture even
when spillovers result. Finds that joint venture may not
benefit the host country, despite spillovers.

Nakamura, M., and J. Xie. 1998. "Nonverifiability, Non
contractibility, and Ownership Determination Models
in Foreign Direct Investment, with an Application to
Foreign Operations in Japan." *International Journal of
Industrial Organization* 16 (5): 571 99. Finds support
for theory that ownership share depends on the impor
tance of the firm's intangible assets and its bargaining
power relative to local partners, using data on foreign
manufacturing operations in Japan.

Raff, Horst, Michael Ryan, and Frank Stähler, 2007.
"Whole versus Shared Ownership of Foreign Affiliates."
Kiel Economics Working Paper 2007 18. Available at
http://opus.zbw kiel.de/volltexte/2007/5684/pdf/EWP
2007 18.pdf. Shows how partial ownership can be used
as a screening device to elicit private information from a
local firm. Application to Japanese firm level data.

HORST RAFF

■ knowledge-capital model of the multinational enterprise

The knowledge-capital model of the multinational enterprise (KC model hereafter) incorporates the theories of horizontal and vertical foreign direct investment (FDI) into a model of the New Trade Theory (see Markusen 2002). The KC model is then used to explore how the location decision of potential multinational firms, their production decisions, the international trade structure (and volume), factor prices, and welfare depend on relative and absolute country size, relative factor endowments, trade barriers, and investment costs.

The term *knowledge-capital model* is derived from the assumption that multinational firms have an ownership advantage compared with other firms due to some knowledge asset, such as patents, blueprints, procedures, brand names, trademarks, or reputation. Knowledge capital has three properties. First, its usage in plants is cheap to separate geographically from its creation in headquarters (fragmentation). Second, knowledge-capital creation is skilled labor intensive (skilled labor intensity). Third, once produced in the headquarters, it can be used in multiple plants within the firm (joint input). The KC model regards FDI as a flow of knowledge in the form of managerial and engineering services, financial services, reputation, and trademarks across borders.

From the central features of knowledge capital derive several location advantages that countries must have to attract FDI. Since the headquarter activity is skilled labor intensive while production is relatively unskilled labor intensive, factor costs are minimized by locating headquarters in relatively skilled-labor-abundant countries and plants in unskilled-labor-abundant countries, motivating vertical FDI. However, separation of production from the headquarters will require trade of goods across borders to serve the home market. To avoid trade costs, producing locally in several locations (horizontal FDI) may be worthwhile. Since there are plant-level fixed costs in addition to firm-level fixed costs, host country markets need to be sufficiently large because otherwise the profits from saving trade costs would not cover the additional plant-fixed costs of local production compared with exporting. If a foreign market is small, it may be optimal for the firm to abstain from production abroad and export from home instead.

Due to knowledge capital, firms favor corporate control of foreign operations over outsourcing. With international outsourcing, the knowledge capital may dissipate faster to competitors (internalization advantage).

Model Description In the standard KC model, there are two countries, two factors of production (skilled and unskilled labor), and two homogeneous goods. Countries differ in their factor endowments. Production of one good is perfectly competitive and subject to constant returns to scale (competitive sector). FDI can occur only in the second sector, which operates at increasing returns to scale, using a three-stage production process (increasing-returns sector). In the first stage, a firm must undertake some headquarter services such as R&D, management, accounting or marketing activities, and firm-level

fixed costs accrue. In the second stage, plant-level fixed costs are incurred. In the final stage, firms compete in an oligopolistic goods market, choosing simultaneously supply (Cournot competition). Markets are segmented and part of the export value is wasted (iceberg trade costs) when shipping goods of the increasing-returns sector across borders.

Assuming headquarter activities are more skilled labor intensive relative to integrated plant and final production in the increasing-returns sector, which, in turn, is more skilled labor intensive than production in the competitive sector, three firm types can emerge:

- *Horizontal* FDI firms have their headquarter activities tied to the home plant and duplicate the domestic production plant in the host country. These firms sell their entire production locally.
- *Vertical* FDI firms locate skilled-labor-intensive headquarter services in the skilled-labor-abundant home country and relatively unskilled-labor-intensive production activity in the unskilled-labor-abundant host country. Foreign affiliates of the vertical type export (part of) their production to the home country.
- *National* firms solely produce in the home country and serve foreign markets by exports.

Firm Location, Foreign Affiliate Production, and International Trade The model is too complex to be solved analytically and must be explored numerically. A standard display of such numerical analysis uses three-dimensional diagrams based on the Edgeworth box (see figure 1). Given a fixed world endowment in both production factors, the bottom of the box depicts the two countries' shares in the world endowments' skilled and unskilled labor, S and U, respectively. At the left corner is the origin of the first country denoted i where its share both in skilled and unskilled labor is zero. Conversely, the origin of the other country j is at the right corner.

First consider the effects of varying the relative size of the two countries. Along the diagonal connecting the two origins the relative factor endowments are identical in both countries. Beginning along this diagonal with the i country very small in terms of endowments, the prevalent firm types are vertical FDI firms with headquarters in i and national firms with headquarters in j. National firms of country j benefit from their large home market, saving more on foregone plant-fixed costs by exporting than losing from variable trade costs, because their foreign market is small. Vertical FDI from country i benefits also from the large market of country j, minimizing trade costs. At this stage country i firms are multinational, but they are few. There are simply not enough endowments in skilled labor to support many headquarters. Hence aggregate sales of affiliates from country i are small. Affiliate sales start rising with a proportional increase in country i's world endowment shares of skilled and unskilled labor. Horizontal FDI firms with headquarters in i enter, as the home market size grows sufficiently for operating profits to cover the fixed cost.

At still larger relative size of country i, the national firms of country j begin to turn into horizontal firms. Since j-firms then gain competitiveness on the i-market by saving trade cost, they push some i-horizontal firms out of the market, which causes a dip in aggregate affiliate sales until only horizontal firms are left.

Increasing country i's relative size further, there is a continuous rise in affiliate sales, because the larger endowments of i support continuously more headquarters of horizontal FDI firms and correspondingly more foreign affiliates. A peak in country i's affiliate sales are attained when it is slightly larger than country j.

At still a further increase in country i's relative size, its aggregate affiliate sales tend toward zero rapidly, because the foreign market size shrinks and country i's horizontal FDI firms are gradually replaced by national firms without affiliates.

Along the ray connecting the origins, the larger country exports the increasing returns and imports the competitive good because of its comparative advantage in the increasing-returns sector.

Next, the relative endowment shares are varied, keeping income of both countries and the world
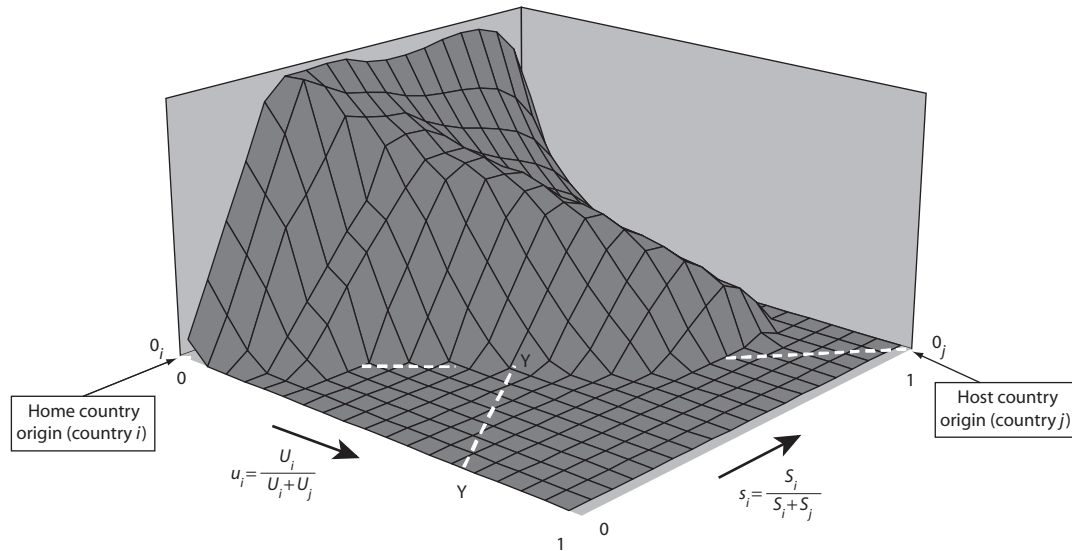
Figure 1
Simulated affiliate sales based on the KC model from Markusen (2002)

endowment equal, which is approximately true for endowment combinations along the line denoted YY in figure 1. Affiliate sales of firms with headquarters in country i have an inverse U-shape along this line.

When countries are identical and trade costs sufficiently high or world endowments sufficiently large, then horizontal FDI firms prevail, and there are equally many in each country. As country i's skilled-endowment share rises relative to country j's, production plants remain equally spread. More and more headquarters relocate to country i, however, since headquarters are skilled labor intensive, absorbing the excessive skilled-labor supply there.

The peak in aggregate affiliate sales of firms with headquarters in country i occurs when the i country is moderately skilled labor abundant. Then country i is residence to a maximum number of headquarters from horizontal-FDI firms while the number of j-country horizontal-FDI firms has shrunk to zero.

Further increasing relative skilled labor abundance of country i, horizontal firms begin to switch gradually into national firms, and country i begins to export the increasing returns good in exchange for the competitive good. Since all world headquarters are already located in country i and still dispropor-

tionately much skilled labor is left in country i, the second most intensive activity production of the increasing returns good is located there, too. But this makes up national firms that do not have foreign affiliate sales. The more horizontal firms are replaced by national firms, the less aggregate affiliate sales country i firms will have.

When almost all skilled labor of the world is concentrated in country i, horizontal firms vanish and vertical FDI firms emerge next to the national firms all with headquarters in country i. Vertical FDI firms of country i have only headquarter activities there, which absorbs the largest possible amount of skilled relative to unskilled labor.

Overall, the peak of affiliate sales due to vertical FDI occurs when the home country is small and relatively skilled labor abundant. A peak of horizontal affiliate sales occurs when both countries are similar in size and relative factor endowments. National firms are prevalent when the home country is large and skilled labor abundant.

FDI Liberalization, Factor Prices, and Welfare
How does the factor income distribution and countries' welfare change if FDI is inhibited initially, and then free to exist? Such FDI liberalization

increases the skilled-wage premium (the ratio of the wage paid to skilled labor relative to that paid to unskilled labor) of the skilled-labor-abundant country, because the specialization in headquarters either of the horizontal or vertical type shifts labor demand toward skilled labor. If relative factor endowments are equal across the two countries, the skilled-wage premium will rise nevertheless whenever horizontal FDI emerges (at sufficiently similar country sizes), because horizontal FDI incurs fixed costs for two plants rather than one, and fixed costs will be intensive in skilled labor. The relative skilled-wage premium even rises in the unskilled-labor-abundant country when the relative abundance is very strong and relative country size similar. The reason is that production plants of more-skilled-labor-intensive vertical FDI from the other country begin to replace firms of the unskilled-labor-intensive competitive sector, shifting relative labor demand toward skilled labor.

Instead, when a country is very small and unskilled labor abundant or very large and skilled labor abundant, FDI is unattractive, and neither firm types nor relative factor prices change. Only a large unskilled-labor-abundant country will experience a fall in the skilled-wage premium, as national firms are replaced by affiliates of foreign vertical FDI firms, substituting skilled-labor-intensive headquarters for less-skilled-labor-intensive production activities.

Turning to welfare, countries that lose from investment liberalization are primarily large, or somewhat smaller, and very skilled labor abundant. The welfare loss arises from relocating production of the increasing-returns good to the foreign country while agglomerating the headquarter activities at home. In this case the increasing-returns good needs to be imported at trade costs and the competition on the domestic market between local producers and importers is less fierce, increasing the profit margins. Both the trade costs and the increase in the profit margin contribute to the rise in the relative price of the increasing-returns good, which decreases consumer rent and welfare.

A welfare gain for both countries arises, for example, if both countries are similar in size and relative skill endowments. Both countries gain then from the savings of trade costs and lower product prices through horizontal FDI, which must overcompensate for the additional fixed costs. Otherwise firms would not choose horizontal FDI.

Empirical Evidence The empirical evidence on the KC model is mixed. Pooling inward and outward FDI data for the United States, Carr, Markusen, and Maskus (2001) provide strong support for the KC model, finding that affiliate sales increase in the total income of the host and home countries, in skill differences, and in similarities of country size. In addition, they find that home countries, which are simultaneously skilled labor abundant and small, have higher affiliate sales.

However, Markusen and Maskus (2001) find a negative relation between affiliate sales and skilled labor abundance in the home country when investigating outward U.S. FDI only. Moreover, Markusen and Maskus (2002) cannot prove the KC model favors a horizontal model where vertical FDI is ruled out by assumption. Blonigen, Davies, and Head (2003) first reestimate the U.S. data separately on observations with negative and positive skill differences, and then apply the absolute values of differences in skill endowments and size in pooled regressions. Finding that increasing skill differences whether positive or negative decrease affiliate sales, they conclude this pattern consistent only with horizontal FDI and not with vertical FDI, hence rejecting the KC model. Markusen and Maskus (2003) in their reply call this finding inconsistent with any model.

Recently, some evidence in favor of vertical FDI in the KC model has been provided, however. Braconier, Norbäck, and Urban (2005) collect data on affiliate sales from home countries that are small and skilled labor abundant and add them to the U.S. data to have more observations where vertical FDI is expected. Moreover, they define the relative skill variable not as absolute but as relative distance, which puts more weight on large differences. Then they obtain in their estimates the typical surface of the KC model.

Extension Two stylized facts are not accounted for by the standard model. Bilateral affiliate sales are

positively correlated with FDI stocks, and there are both more bilateral affiliate sales and exports the more similar countries are in size. These two stylized facts are explained by Bergstrand and Egger (2007), assuming that plant (headquarter) fixed costs consist of physical capital from the home country (skilled labor) and that there are three instead of two countries.

Suppose there were only horizontal firms when countries are symmetric in endowments, then demand for physical capital would be at a maximum, raising the rental rate of capital and fixed cost. Since national firms have one plant less, they save on fixed costs at the expense of larger trade cost. Hence a mix of national and horizontal firms will emerge such that both types are equally well off. Bergstrand and Egger show further that the addition of a third country helps explain that both affiliate and export sales rise as two countries become more similar.

In sum, the KC model provides a rich description of the types of activities undertaken by multinational enterprises, including vertical and horizontal FDI. The degree of empirical support for the KC model is undergoing debate, with an extension to include a third country currently showing the most promise.

See also factor endowments and foreign direct investment; fixed costs and foreign direct investment; footloose production; foreign direct investment (FDI); location theory; market access; market size and foreign direct investment; outsourcing/offshoring; proximity-concentration hypothesis; trade costs and foreign direct investment; vertical versus horizontal foreign direct investment

FURTHER READING

Bergstrand, Jeffrey H., and Peter Egger. 2007. "A Knowledge and Physical Capital Model of International Trade Flows, Foreign Direct Investment, and Multinational Enterprises." *Journal of International Economics* 73 (2): 278 308. Important extension of KC model to three countries.

Blonigen, Bruce, Ron Davies, and Keith Head. 2003. "Estimating the Knowledge Capital Model of the Multinational Enterprise: Comment." *American Economic Review* 93 (3): 980 94. Comment on the first estimation of the KC model.

Braconier, Henrik, Pehr Johan Norbäck, and Dieter Urban. 2005. "Reconciling the Evidence on the Knowledge Capital Model." *Review of International Economics* 13 (4): 770 86. Empirical evidence in favor of the KC model.

Carr, David, James R. Markusen, and Keith E. Maskus. 2001. "Estimating the Knowledge Capital Model of the Multinational Enterprise." *American Economic Review* 91 (3): 693 708. First estimation of the KC model.

Markusen, James R. 2002. *Multinational Firms and the Theory of International Trade*. Cambridge, MA: The MIT Press. Seminal publication including many of Markusen's models of FDI.

Markusen, James R., and Keith E. Maskus. 2001. "Multinational Firms: Reconciling Theory and Evidence." In *Topics in Empirical International Economics: A Festschrift in Honor of Robert E. Lipsey*, edited by Magnus Blomstrom and Linda Goldberg. Chicago: Chicago University Press, 71 95. Extension to the first estimation of the KC model.

. 2002. "Discriminating among Alternative Theories of the Multinational Enterprise." *Review of International Economics* 10 (4): 694 707. Mixed evidence on the KC model when discriminating it against the horizontal or vertical FDI model.

. 2003. "Estimating the Knowledge Capital Model of the Multinational Enterprise: Reply." *American Economic Review* 93 (3): 995 1001. Enlightening explanations on how to estimate the KC model and how to discriminate it against alternatives.

DIETER M. URBAN

■ **Kyoto Protocol**

See multilateral environmental agreements

■ labor standards

Cross-country variations in labor market practices and protections can have implications for the pattern and terms of trade. Weak labor protections that augment the supply of unskilled labor, such as forced labor and child labor, can increase the supply of unskilled-labor-intensive exports on world markets, thereby depressing wages of unskilled workers worldwide. Inhumane labor practices may also generate negative reactions from consumers and investors globally. The external effect of one country's labor practices on its trade partners provides a basis for international coordination of labor market regulations. Internationally coordinated labor practices are commonly referred to as *international labor standards*.

Proponents of international coordination of labor standards emphasize the consequences that trade with low-wage countries will have for workers in industrialized countries. In the absence of international coordination, there may be a *race to the bottom* in labor protections. Opponents of coordination argue that international labor standards are a form of protectionism.

International coordination of labor market regulations was first advanced by Switzerland in 1881. A subsequent series of international conferences over the following 32 years produced agreements prohibiting international trade in white phosphorous matches and limits on night work for women.

Currently, labor standards in international trade agreements are limited. Article XX(e) of the World Trade Organization (WTO) Charter permits members to refuse imports of goods produced with prison labor. Broader labor standards have been incorporated into regional trading agreements such as the North American Free Trade Agreement. Further coordination of labor standards within the General Agreement on Tariffs and Trade (GATT), and subsequently the WTO, was proposed by the European Parliament in 1983 and 1994, by the U.S. government in 1986, and at every WTO Ministerial between 1996 and 2001. The Singapore Ministerial Declaration (December 1996), however, while acknowledging the importance of international labor standards, identified the International Labor Organization (ILO) as the competent body to establish and monitor labor standards.

Labor standards commonly fall into two broad groups. *Process*, or *core*, labor standards regulate the basic function of labor markets and are considered by many to be basic human rights. Core labor standards, as identified by the ILO, are (1) rights to free association and collective bargaining, (2) prohibition against forced labor, (3) abolition of exploitative child labor, and (4) elimination of discrimination in employment. *Outcome* standards place limitations on aspects of the labor contract such as hours worked and wages.

Labor Standards and International Trade Negotiations The link between domestic labor regulations and the international terms of trade has implications for international trade negotiations. A labor-scarce country that makes tariff concessions granting market access to a labor-abundant trade partner in a round of international trade negotiations can in effect offset those concessions by easing costly

domestic labor standards that adversely affect import-competing producers. The *ex post* reduction in domestic labor protections is referred to as a *race to the bottom.*

In the absence of a mechanism to restrict *ex post* changes in domestic regulations, the labor-scarce country cannot guarantee the promised market access. For this reason, efficient bargaining requires that tariffs and domestic policy be negotiated simultaneously.

More generally, a country's domestic labor practices generate an external effect on its trade partners through the terms of trade. Labor-using standards, such as a prohibition against prison labor, reduce the supply of labor relative to capital on world markets. As a consequence, the price of labor-intensive goods rises and the price of capital-intensive goods falls on the world market. Thus, labor protections improve the terms of trade for a labor-abundant country but worsen the terms of trade for a labor-scarce country. The existence of such an external effect creates an incentive for labor-abundant countries to overregulate their labor markets (from a global efficiency perspective) while labor-scarce countries have an incentive to underregulate their labor markets.

Existing provisions of the GATT have been proposed to coordinate labor practices internationally; these include anti-dumping measures (Article VI), the prohibition against export subsidies (Article XVI), the opt-out provision (Article XXXV), the Trade Policy Review Mechanism, and the Nullification and Impairment clause (Article XXIII). Of these, only Article XXIII has been found to provide a legal basis for regulating domestic policies within the WTO. Negotiations within the WTO are understood to provide each country with a certain level of access to its trade partners' markets. Any acts by the importer that nullify or impair the expected access can form the basis of a complaint and trigger renegotiation of the tariff agreement. In the context of the labor standards debate, a labor-scarce importer that relaxes its labor protections in an attempt to limit the import exposure created by its tariff reductions would then be required to make additional tariff reductions as compensation to a labor-abundant exporter. The exercise of the Nullification and Impairment clause thus internalizes the external effect of domestic labor standards policies.

Use of Article XXIII has not been pursued in the WTO for the purpose of coordinating labor practices. Rather, two other options were advocated in the 1990s to manage the interrelationship between border controls and domestic policies. Some delegates sought to extend the General Exceptions provisions of Article XX to include a range of labor practices (beyond prison labor). A social clause was also proposed that would harmonize labor standards among members of the WTO. Members were unable to agree, however, on a common set of universal standards that would be enforced within the WTO. At the conclusion of the 1996 Singapore ministerial, the task of establishing and monitoring labor standards was delegated to the ILO, and WTO members committed themselves to cooperate with the ILO on these issues.

The ILO was created in 1919 as part of the League of Nations and became part of the United Nations after 1946. The ILO promulgates conventions and recommendations concerning labor practices that member governments may or may not choose to ratify. Compliance is voluntary but the ILO provides technical assistance on improving labor practice.

Enforcement and Issue Linkage Given the lack of enforcement power within the ILO, it has been recommended that some enforcement power from the WTO be transferred to the ILO. Under linkage, a WTO member, adversely affected by noncompliance with ratified ILO Conventions by its trade partner, could make a complaint under Article XXIII that its expected market access has been nullified or impaired. Some compensation in the form of other tariff concessions could then be required.

Critics of this approach contend that simultaneous negotiations of trade and labor issues will produce stricter labor standards but less trade liberalization than would be otherwise forthcoming from two separate agreements. Issue linkage may also increase enforcement power, however. The impact of coordination depends on how countries calculate the payoffs from cooperation on the two issues. If a

failure to reach a cooperative agreement on one issue increases the value of cooperation on the other issue, these two coordination objectives are *substitutes*. In this case, countries will find it easier to sustain cooperation on either issue if negotiations on both issues are linked in a single agreement. A country considering defection on labor practices, for example, will see cooperation collapse on labor and trade if cooperation is linked in a single agreement. The consequent increase in the cost of defection achieved by linking issues will strengthen the capacity of countries to negotiate self-enforcing agreements on both trade and labor.

Market Mechanisms Market-based mechanisms have also been proposed for improving labor practices. Product labels characterizing conditions of work in the production process offer consumers an opportunity to pay a premium for goods produced under humane working conditions. Market efficiency will be achieved with a credible product label if consumers have a *private* distaste for poor working conditions.

By contrast, if the external effect of working conditions on consumers is a *public* good, product labels may not correct the market failure. In the case of child labor, firms seeking a *child labor free* label will be required to discharge underage employees. If child employment in the export sector is the first-best option for the child, termination of employment will force the child to the second-best choice. Child welfare will only improve if (1) product labeling is sufficiently pervasive that it raises the wages of adult labor above the level at which families withdraw children from the labor force, or (2) some portion of the premium paid by consumers is diverted to a fund used to provide welfare services to former child workers.

Corporate codes of conduct relating to the general conditions of work for adult employees have greater potential to improve working conditions. First, codes may simply require factories to employ more humane, though more costly, means of production in order to satisfy the demands of consumers and stockholders. Second, compliance officers may impart knowledge concerning labor management practices that are both more humane and more efficient. For example, factory managers may not fully appreciate the role harsh treatment and low wages play in increasing manpower turnover and retaining investment in skill trainings. Production inefficiencies such as poor line organization may have the effect of increasing overtime, thereby reducing productivity and wages. Third, corporate compliance officers that enforce codes requiring free association and collective bargaining may also inhibit monopsonistic employment practices, such as exploiting the lack of market experience of young female employees.

At the end of 2007, labor standards in the world economy are limited to voluntary initiatives and regional trading arrangements. Individual countries establish their own labor protections within the confines of a set of global ethical and domestic political considerations. Concern with the impact of trade on labor ebbed during the early part of the 21st century. Continuing integration of global labor markets may increase the pressure to implement labor protections, however.

See also child labor; comparative advantage; General Agreement on Tariffs and Trade (GATT); International Labor Organization; terms of trade; trade and wages; World Trade Organization

FURTHER READING

Basu, Kaushik, Henrik Horn, Lisa Roman, and Judith Shapiro, eds. *International Labor Standards*. Malden, MA: Blackwell, 2003. Four papers covering the theory and history of labor standards, an application to child labor, and a discussion of the role of labor standards in international organizations.

Brown, Drusilla K. 2001. "Labor Standards: Where Do They Belong on the International Trade Agenda?" *Journal of Economic Perspectives* 15 (3): 89 112. A concise discussion of the basic issues relating to labor standards and trade policy.

DRUSILLA K. BROWN

■ Latin American debt crisis

The Latin American debt crisis officially began in August 1982 with Mexico's announcement that it

could no longer meet its interest payment obligations. External debt in Latin America had quadrupled from a 1975 level of $75 billion to more than $315 billion by 1983, which was about 50 percent of gross domestic product for that region. Variable interest rate loans meant that debt service (interest payments and the repayment of principle) grew even faster, rising to $66 billion by 1982 from $12 billion in 1975 (Gruppen 1986). The crisis was set in motion in the early 1970s when private bank lending to Latin America began to expand rapidly, eventually overtaking international development bank lending as the primary source of external funding for many rapidly growing Latin American countries.

This excessive private lending has roots in the oil price dynamics of the early 1970s. In 1973, the Organization of the Petroleum Exporting Countries (OPEC) cut production for the first time, resulting in a fourfold increase in the price of oil worldwide. This phenomenon, combined with the fact that oil bills were paid almost exclusively in dollars, resulted in large inflows of dollars into oil-producing countries. For the most part, oil-rich OPEC countries lacked an industrial manufacturing infrastructure that could absorb significant new domestic investment in the short term. The only choice for OPEC countries was to deposit dollars in what was by then a relatively well-established Eurodollar market. This consisted mostly of U.S. and Japanese banks with branches in Europe that accepted dollar deposits. These massive deposits quickly became known as petrodollars. The stage was set for a massive lending spree from private banks to newly industrializing countries, especially in Latin America. Most scholars agree that there was significant overlending during the period from 1973 to 1982. The important question is why.

From Boom to Bust On the supply side of the loanable funds market, the answer is relatively clear. Eurodollar deposits dramatically increased banks' desire to lend. Banks saw in oil-exporting less-developed countries (LDCs) an opportunity for big profits. Given that traditional business-sector demand for loans in developed countries (DCs) had all but dried up due to the supply shock induced recession (a sudden inward shift in the aggregate supply

curve caused by the dramatic increase in oil-related manufacturing costs), the Latin American loan market served as an attractive alternative. Moreover, authoritarian governments were standing at the ready to accept and guarantee repayment of private bank loans. This was a bonus for banks, which worked on the principle that sovereign loans (that is, loans to governments) carried very little risk of default compared with private sector loans.

On the demand side of the loanable funds market, Latin American countries exhibited considerable eagerness to borrow. One reason for this eagerness stemmed from the exigencies of state-led growth. By the late 1960s, Latin American newly industrialized economies (NIEs) faced the end of what scholars have referred to as "the easy phase" of import-substituting industrialization (ISI). The import-substituting strategy was first employed by most Latin American countries in the 1930s in an effort to break dependence on primary exports by producing previously imported consumer goods for the domestic market. By the late 1960s, continued growth under the ISI model meant producing more capital intensive consumer goods. This meant that if state-led growth was going to continue at comparable rates, new investments in heavy manufactured goods (i.e., cars and other consumer durables) would have to take place (Krueger 1993). Moreover, the majority of Latin American NIEs were governed by bureaucratic authoritarian regimes whose legitimacy hinged on the continued ability to deliver rapid industrial growth.

Some scholars have argued that Asian countries were better equipped to deal with the external shocks of the 1970s because they had already begun to shift their policies from import substitution to an export-oriented economy (Sachs 1985). In Latin American countries, the choice to deepen ISI rather than abandon it for an export-led approach was reinforced by the fact that the 1973 oil crisis caused severe recession in most developed countries, thus drying up the potential export market for Latin American NIEs. One consequence of the push toward heavy manufactured products in Latin America was that non-oil exporters became desperate for funds to

cover increasing production costs and growing balance of payments deficits.

Certainly, these countries could have turned to their traditional source of balance of payments financing, the International Monetary Fund (IMF), but the fact that they could borrow from private banks allowed them to avoid IMF conditionality, which had always been a point of political discontent (Bird 1996). But even the oil-exporting LDCs in Latin America, Mexico and Venezuela in particular, saw the availability of private bank credit as an opportunity for industrialization backed by state leadership.

The most obvious incentive for Latin America to borrow during the 1970s grew out of financial market conditions. Inflation had skyrocketed after the 1973 oil price hikes while nominal interest rates had leveled off and even fallen in some cases due to the credit glut. Together these phenomena resulted in extremely low and even negative real interest rates. For a time, banks were paying Latin American governments to take their money (Sampson 1981). Ultimately, the buildup of excessive debt in the 1970s grew out of a coincidence of political and economic interests among Latin American NIEs and private banks.

The stage had been set for a major crisis when another oil crisis hit in 1979. This time oil prices did not rise by nearly as much as they had in 1973. The effects of 1973 had been stagflation: severe recession coupled with accelerating inflation. This time, the crisis led to deepened recession but very little inflation. Part of the reason was that the Federal Reserve Board, under Paul Volcker's conservative monetary leadership, opted to raise interest rates to fight inflation despite continuing recessionary conditions. As a result, real interest rates on existing variable rate loans began to rise rapidly. For non-oil-producing LDCs, the combination of high oil prices, recession, and skyrocketing loan commitments spelled disaster. But even Mexico, the model debtor, had come to discover that its oil fields were much less extensive than originally predicted, while it still faced all of the aforementioned challenges. Meanwhile, lenders became increasingly hesitant to make new loans as the

credit glut conditions of the early 1970s had largely disappeared. Not surprisingly, although it seemed to take the major players by surprise, Mexico announced in August 1982 that it could not meet even its interest payments on its massive debt obligations, thus triggering the debt crisis.

Crisis Resolution and Reform In response to the crisis, the Baker Plan (named for then U.S. treasury secretary James Baker who proposed the plan) focused on rescheduling payments and arranged a $12 billion rescue package funded primarily by the United States and the IMF to Mexico. The Mexico loans were part of a larger $29 billion proposal for the Latin American and several other debtor nations (Bogdanowicz-Bindert 1986). The plan did not get banks to resume lending, although banks did eventually discount, forgive, and/or reschedule existing debt. Ultimately, the crisis underscored a problem of debt dynamics rather than sheer mass. In other words, the Latin American debt crisis points to a breakdown in the cycle of lending and repaying. The amount of debt buildup in the 1970s was by no means unique historically, but the breakdown or interruption of the financial system cycle of lending and repaying was unprecedented (Fishlow 1985). In a panicked response to Mexico's announcement that it could not pay, banks immediately stopped lending to all developing countries, which triggered a massive and unprecedented debt crisis because even relatively sound borrowers could not make payments on existing loans without access to new lines of credit. This dynamic problem of debt buildup and repayment problems has led to considerable debate on the merits of debt relief and debt forgiveness. The Debt Laffer Curve suggests there is a critical debt stock beyond which both the lenders and borrowers lose. That is, as the external debt stock rises, the indebted country will try to produce less (discouragement effect) or intentionally default on the existing debt (sabotage) so the foreign lenders will receive less than full repayment. If the debt stock is already above this level, as it certainly was in 1982, it may be in the interest of the lenders to forgive some of the debt.

In the end, the lasting story behind the Latin American debt crisis revolves around its consequences

for Latin American countries and particularly the most vulnerable segments of the population. Between 1982 and 1984, the net transfer of capital from Latin America to developed countries' banks approached $75 billion, an amount equal to one-quarter of all Latin American exports for the same period. Growth rates and living standards declined in the region for virtually the entire decade of the 1980s, a decade now commonly referred to as "the lost decade" in Latin America.

The Latin American debt crisis had long-lasting political, economic, and social repercussions. Ultimately, the debt crisis led to an abandonment of import substitution industrialization policies and a move toward trade and financial liberalization. Proponents of neoliberalism thus see the debt crisis as a difficult but necessary lesson for policymakers and their constituents about the inefficiency of state interventionist policies. Opponents of neoliberalism view the debt crisis as a tragedy that put Latin American countries in the unenviable position of having to choose between adopting neoliberal policy reform and remaining cut off from global financial markets. Regardless of which view one takes, there is little dispute that the Latin American debt crisis triggered an era of neoliberal reform.

See also bail-ins; banking crisis; capital flows to developing countries; currency crisis; Eurocurrencies; Federal Reserve Board; financial crisis; financial liberalization; International Monetary Fund (IMF); peso problem; petrodollars, recycling of; sovereign risk; Washington consensus

FURTHER READING

Auerbach, Nancy Neiman. 2007. "The Meanings of Neoliberalism." In *Neoliberalism: National and Regional Experiments with Global Ideas*, edited by R. Roy, A. Denzau, and D. North. New York: Routledge, 26 50. Identifies the Latin American debt crisis as the impetus that not only leads to the neoliberal era in Latin America but gives Latin American neoliberalism some of its unique characteristics.

Bird, G. 1996. " The International Monetary Fund and Developing Countries: A Review of the Evidence and Policy Options." *International Organization* 50 (3): 477 511. Using a political economy framework, analyzes the part played by the IMF as a balance of payments financing and economic adjustment organization.

Bogdanowicz Bindert, C. A. 1986. " The Debt Crisis: The Baker Plan Revisited." *Journal of Interamerican Studies and World Affairs* 28 (3): 33 45. Describes the immediate and ongoing effects of the Baker Plan on debt crisis countries.

Fishlow, Albert. 1985. "Lessons from the Past." In *The Politics of International Debt*, edited by Miles Kahler. Ithaca, NY: Cornell University Press, 37 93. Argues that the major difference between the Latin American debt crisis and other historical instances where countries were so heavily indebted that they threatened default is that private lenders stopped lending cold turkey and created a dynamic crisis where the lack of new lending made it impossible to pay off existing loans.

Gruppen, N. 1986. *The Debt Crisis in Latin America*. Stockholm: Institute of Latin American Studies. Detailed overview of the financial picture leading up to and precipitating the Latin American debt crisis.

Krueger, A. O. 1993. " Virtuous and Vicious Circles in Economic Development." *The American Economic Review* 83 (2) (Papers and Proceedings of the 105th Annual Meeting of the American Economic Association): 351 55. Focuses on the political economic interactions surrounding externally versus internally oriented trade policies. It suggests that Pareto optimal policies cannot be identified absent consideration of political responses.

Sachs, Jeffrey. 1985. "External Debt and Macroeconomic Performance in Latin America and East Asia." *Brookings Papers on Economic Activity* 2: 523 64. Provides an excellent overview of the macroeconomic effects of heavy debt burdens in Latin America.

Sampson, A. 1981. *The Money Lenders*. New York: Penguin. Paints a vivid picture of the lending market in the 1970s when private banks literally begged Latin American finance ministers to take the money off their hands.

Wellons, P. A. 1985. " International Debt: The Behavior of Banks in a Politicized Environment." *International Organization* 39 (3): 441 71. One of the few scholarly pieces that focuses on explaining bank behavior within a political context leading up to the debt crisis.

NANCY NEIMAN AUERBACH

■ **law of one price**
See purchasing power parity

■ **learning by doing**
See infant industry argument

■ **lender of last resort**
The classical role of a lender of last resort (LOLR) is to provide solvent participants of the financial system with liquidity in times of crisis. In practice, however, the LOLR function of central banks has been used to help banks in distress. Rules with respect to the LOLR are often ambiguous. In the European Monetary Union (EMU) in particular, the European Central Bank (ECB) as well as national central banks are assigned LOLR responsibility. On the international level, the International Monetary Fund (IMF) is sometimes viewed as a LOLR for governments.

The Financial System Safety Net Banks have a special role in the financial system as the main providers of liquidity. Behind the commercial banks, a central bank uses monetary policy instruments to control the availability of liquidity in an economy. In some countries central banks are responsible for financial supervision, and banking supervision in particular, as well as monetary policy. In other countries, a financial supervisory authority is responsible for the "safety and soundness" of individual banks, while the central bank retains responsibility for the payments system.

As a result of the important role of banks in the payments system and the risk of runs on banks, most countries have developed a "safety net" for banks, in particular. The safety net includes a LOLR function along with deposit insurance, capital requirements, regulation, and supervision. These aspects of the safety net should jointly protect the banking system while providing banks with the appropriate rules and incentives to allocate credit efficiently.

Two Views of the LOLR The classical formulation of the LOLR function goes back to Henry Thornton in 1802 and Walter Bagehot in 1848, and the term goes back even further to Sir Francis Barings, who referred to the Bank of England as the "dernier resort" providing liquidity to banks in times of crisis. Bagehot's classical formulation as interpreted by Humphrey (1989) and Gaspar (2006) includes the following characteristics:

1. The LOLR has the objective of protecting the integrity of the financial system rather than individual institutions;
2. The LOLR supports the central bank's monetary policy objectives;
3. Insolvent institutions should be allowed to fail;
4. LOLR assistance should be provided to solvent, illiquid institutions;
5. LOLR lending should be granted at penalty rates;
6. LOLR lending should be granted only against good collateral; and
7. Conditions for LOLR lending should be announced and well understood before a crisis event.

Although these characteristics lack specificity with respect to, for example, rate setting, acceptable collateral, and operational definitions of liquidity and solvency, they quite clearly identify the LOLR function as an aspect of monetary policy in times of potential or actual liquidity problems in the economy. Thus, in this classical view, the LOLR stands ready to lend to any entity offering collateral and willing to pay the penalty rate. An example of this LOLR function in practice occurred on September 11, 2001. After the terrorist attacks on New York and Washington, the Federal Reserve in the United States pronounced its readiness to supply liquidity "to support the economic and financial system," and the European Central Bank (ECB) also stood ready to "support the normal functioning of the markets" (see Gaspar 2006).

The more common interpretation of the LOLR, as well as the most commonly practiced role, is that the central bank supplies emergency liquidity assistance to specific financial institutions. In this case, the LOLR function becomes part of the crisis management framework for specific institutions in distress.

The requirement that LOLR lending should be reserved for solvent banks facing liquidity problems remains valid under this interpretation. The rationale for this LOLR function is the common view that contagion of one bank's problem to other solvent banks can occur as a result of the problems of evaluating the risk level of individual banks. Thus LOLR activities become oriented toward specific banks.

The difficulty of distinguishing between liquidity and solvency problems in a banking crisis often leads central banks to provide assistance to insolvent banks. Thereby, the LOLR function of central banks can become the source of moral hazard in banking; if emergency assistance can be expected, bank managers' incentives to extend risky loans increase. Furthermore, emergency assistance to insolvent banks causes delays in closing the banks. Thereby, the final costs of a bank failure for creditors and taxpayers may increase.

When a central bank provides liquidity assistance it faces a trade-off between reducing the risk of contagion and reducing the moral hazard problem caused by expectations of aid to insolvent banks.

The EMU System For international banks working within the jurisdictions of several central banks it can be important whether the home or the host country central bank serves as the LOLR. In the EMU, responsibility for monetary policy, which includes the LOLR function, lies clearly with the ECB.

The assignment of responsibility with respect to emergency liquidity assistance for an international bank is less clear. On the face of it, one expects that emergency liquidity would be provided by the central bank in the country where the bank is incorporated. Subsidiaries are often strongly integrated with the parent, however, and therefore indirectly beneficiaries of home country assistance. Assistance to a bank in crisis can arise suddenly and conflicts of interest with respect to burden-sharing are likely. Furthermore, emergency assistance by a LOLR is only one form of assistance to a bank in distress. Thus there is a need to coordinate activities of the central bank, financial supervisors, and fiscal authorities.

In the EU and, therefore, in the EMU, responsibility for financial stability, bank supervision, and crisis management is decentralized to the national level. Thus, LOLR responsibility in the form of emergency liquidity assistance rests with the individual countries, while the classical LOLR function as a provider of liquidity to the financial system as a whole is the responsibility of the ECB.

The ambiguity in the EMU arises because emergency liquidity assistance by a national central bank has monetary implications. Monetary policy can be conducted only by the ECB, however. As pointed out by the European Shadow Financial Services Regulatory Committee (1998) a committee of independent experts analyzing and publicly commenting on regulatory issues "any operation that is undertaken on the national level has EMU-wide monetary repercussions. For example, an interest rate subsidy to a local problem bank may in the end be paid for by other banks in the EMU and their customers." The committee recommended that procedures for emergency liquidity assistance should be established within the EU to resolve this ambiguity.

According to some commentators a degree of "constructive ambiguity" is desirable in crisis management on the grounds that it can reduce the moral hazard problem discussed earlier. Indeed, ambiguity is what EU law provides with respect to the division of LOLR responsibility between the ECB and national central banks in the Euro system. It may take a first pan-European crisis to bring some clarity to the division of LOLR responsibilities within the EU.

The IMF as International Lender of Last Resort The International Monetary Fund's (IMF's) role in helping Mexico during the Tequila crisis in 1995 has been widely blamed for contributing to the Asian crisis in 1997 by causing expectations among lenders that loans to sovereign countries are safe. The IMF does not protect banks directly but through its assistance to countries like Mexico in balance of payments crises or when the threat of crisis is substantial, it may provide indirect protection for banks that have lent to the government in a crisis country.

The proper role of the IMF in sovereign crises has been the subject of debate after the Mexican and

Asian crises. Prominent economists like Meltzer (1998) have suggested that the IMF should limit its role in crises to serve as a LOLR in the classical sense. In other words, the IMF should provide loans to bridge a crisis only in the case of pure liquidity crisis. In such a crisis a country's fundamental ability to repay loans is not impaired. Instead, the crisis could arise because a country has run out of foreign exchange reserves, and liquidity is not forthcoming in international financial markets as a result of a lack of information about a country's prospects. In this situation, based on its superior access to information about countries' prospects, the IMF could step in.

Focusing the IMF's role in country crises as an LOLR when there are liquidity crises could require financial resources beyond the capability of the IMF. Many observers argue, however, that the IMF's involvement, even on a small scale, will increase the willingness of private lenders to extend loans to countries in crisis, since the IMF is considered particularly well informed about economic conditions in member countries.

See also asymmetric information; banking crisis; currency crisis; deposit insurance; European Central Bank; European Monetary Union; Federal Reserve Board; international financial architecture; international liquidity; International Monetary Fund (IMF); International Monetary Fund conditionality; international reserves; money supply

FURTHER READING

Eichengreen, Barry. 2004. *Capital Flows and Crises*. Cambridge, MA: MIT Press. Provides analyses of the origin of sovereign crises and the appropriate response of international financial institutions.

European Shadow Financial Regulatory Committee. 1998. "EMU, the ECB, and Financial Supervision." Statement No. 2 (October 1). Downloadable at http://www.ceps.be. Addresses the role of the ECB in financial supervision and crisis resolution.

Gaspar, Vitor. 2006. "Bagehot and Coase Meet the Single European Market." Conference Paper, Federal Reserve Bank of Chicago (October 6). Argues that the European Central Bank is suitable as the Lender of Last Resort in the classical sense.

Goodhart, Charles A. E. 1995. *The Central Bank and the Financial System*. Cambridge, MA: MIT Press. Provides a comprehensive analytical treatment of the different roles of the central bank in the economy

Goodhart, Charles A. E., and Haizhou Huang. 2005. "The Lender of Last Resort." *Journal of Banking and Finance* 29: 1059 82. Models the trade off between costs in terms of risk taking incentives and benefits of reduced contagion.

Humphrey, Thomas M. 1989. "Lender of Last Resort: The Concept in History." *Federal Reserve Bank of Richmond Economic Review* 75 (March/April): 8 16. Reviews the history of the lender of last resort function of central banks and provides an economic analysis of its historical development.

Lastra, Rosa M. 1999. "Lenders of Last Resort: An International Perspective." *The International Comparative Law Quarterly* 48: 340 65. Discusses alternative views, as well as benefits and costs, of the lender of last resort and compares its role in Europe, the United States, and Asia.

———. 2006. *Legal Foundations of International Monetary Stability*. Oxford: Oxford University Press. Provides a legal perspective on central banking, including responsibility and contents of the lender of last resort function on the national, the EU, and the international level.

Meltzer, Alan. 1998. "The Asian Problem of the IMF." Testimony prepared for the Joint Economic Committee, U.S. Congress (24 February), Washington, DC. Presents a critique of the IMF's activities in the Asian crisis.

CLAS WIHLBORG

■ linkages, backward and forward

Linkages are input-output relationships between firms or industrial sectors in the same economy. A firm purchasing inputs from a local supplier is an example of a backward linkage, while a firm selling intermediate inputs to another firm creates a forward linkage. The importance of such linkages with the local economy for economic development has long been recognized in the economics literature. For example, Marshall (1920) argues that input-output relationships between firms are one of the advantages of localized industry, since "subsidiary trades grow

up in the neighborhood, supplying [the firm] with implements and material, organizing its traffic, and in many ways conducing to the economy of its material" (271).

In an early view, Hirschman (1958) discusses the importance of linkages between sectors in an economy in the context of his strategy of unbalanced growth for developing countries. At the heart of this strategy is the suggestion that developing countries should aim at generating imbalances in demand and supply in order to achieve continuing growth. An important part of the unbalanced growth strategy is the potential to develop linkages between sectors in which a leading sector, through linkages with a follower sector, may foster the development of the latter industry.

The issue of local linkages has regained importance in the last decades of the 20th century due to the increasing globalization of the world economy, where many inputs and outputs can be internationally traded, and therefore, inputs can easily be sourced locally or abroad. Given the increasing significance of foreign direct investment and multinational companies in many developed and developing countries alike, a particular focus of interest has been on interfirm linkages between foreign multinationals and indigenous firms, which to some extent echo the role of linkages in Hirschman's earlier work.

Whereas in Hirschman's concept of unbalanced growth "leading sectors" (which have the highest potential for linkages creation) induce growth in other sectors, the more recent view on multinationals largely sees these as "leading firms" inducing growth in local enterprises. Hence one can think of the former concept as intersectoral linkages, while the latter are interfirm linkages. Essentially, the interfirm linkage is at the heart of the Hirschman-style intersectoral linkage, since a sector is composed of a number of firms. The concept of interfirm linkages is a much wider and richer concept, however, and the intersectoral linkage leading to economic development is only one of many effects of interfirm linkages.

Intersectoral Linkages Hirschman's (1958) theory of unbalanced growth suggests that development policy should focus on promoting the growth of "leading sectors," which would result in growth being transferred from these key sectors to the followers. The way of linking key sectors and induced sectors is the development of backward and forward linkages between these sectors. In his words: "The input-provision, derived demand, or *backward linkage effects* . . . will induce attempts to supply through domestic production the inputs needed in that activity," and "the output-utilization or *forward linkage effects* . . . will induce attempts to utilize outputs as inputs in some new activities" (100).

Clearly, Hirschman has linkages between different sectors in mind when he defines these concepts. Also, linkages are seen as a causal (ex ante) concept, that is, linkages from the leading sector A to the following sector B allow sector B to develop; in effect, sector B would not have come into existence without the linkage from sector A. This point indicates the importance of intersectoral linkages and the benefit of these linkages for the development of other sectors and for the economy as a whole.

Intersectoral linkages are generally measured using input-output (I-O) tables, that is, tables of all inputs and outputs of an economy's industries, including intermediate transactions, primary inputs, and sales to final users. As a crude measure, backward linkages can be estimated as the proportion of an industry's output that are purchases from other industries, while forward linkages are the proportion of an industry's total output that is not directed to final demand but rather is an input used by other industries. Based on this idea several more refined indexes based on input-output tables have been suggested in the literature to measure linkages (e.g., Yotopoulos and Nugent 1976; Lall 1980).

These ratios measure only the degree of interdependence of different sectors, to what degree a sector uses inputs from other sectors, and not necessarily the true, induced, linkage effects, however. Although the true linkage is a causal concept, that a sector develops strictly as a result of a linkage to a leading sector only, one could also find sectors with high ratios that are not due to linkage effects but are only due to sectoral interdependence. In this case, the sector did not de-

velop as a result of linkages to another sector but rather *in tandem* with it.

Interfirm Linkages In the more recent literature the focus on local linkages has shifted away from being concerned with sectors to individual firms. Hence, research has attempted to determine what the implications of such linkages between firms are for the development of the local economy. Another focus is the measurement of linkages, in order to determine whether individual firms create backward linkages specifically through the purchase of inputs in the local market or whether they create forward linkages through selling output as intermediate inputs to other local firms.

One reason for the interest in interfirm linkages is the increasing globalization of production and the importance of multinational companies and foreign direct investment for host countries. A crucial question is whether foreign multinationals create linkages with the host economy or whether they operate in so-called enclave sectors with no links to the domestic economy that surrounds them. This issue is important in a world economy where many inputs and outputs can be internationally traded. In such a context it seems important to establish whether firms create backward or forward linkages with the local economy or whether they import and export most or all of their inputs and output. Of course, the answer to this question may change over time, as multinationals may initially source inputs from suppliers in their home country but over time develop relationships with local suppliers.

Furthermore, backward interfirm linkages become important when analyzing local content requirements in international trade agreements. Regional trade blocks, such as the European Union or the parties to the North American Free Trade Agreement, frequently impose minimum requirements of domestic inputs that foreign firms must use in their production. An analysis of backward interfirm linkages shows the extent of local buying and may indicate whether or not local content requirements are met by foreign multinationals in the host economy.

In a world with imperfect competition and economies of scale, interfirm linkages between mul-

tinationals and domestic firms can have positive effects on firms through the emergence of externalities. This issue has attracted considerable attention in the economics literature, and various positive effects of interfirm linkages have been identified. Most directly, interfirm linkages can have effects on the creation of secondary (or indirect) employment generated in supplier firms. If firms are linked through seller-supplier linkages, sellers will have a role to play for employment in the supplier firms.

Theoretical models and empirical research furthermore show some more indirect effects of linkages between multinationals and domestic firms. For example, linkages may encourage the establishment and growth of local companies in the same or different sectors in the host country (Markusen and Venables 1999). Backward linkages, created through the purchase of intermediate goods by final good producing multinationals, can reduce costs for local intermediate good suppliers by increasing the scale of their production. These reduced costs imply, in turn, that final good producers have to incur fewer costs for intermediate products, which is the forward linkage. In such a setting, if multinationals increase demand for intermediate products through backward linkages, indigenous intermediate good suppliers, in turn, can establish forward linkages with indigenous final good producers. Thus backward linkages by multinationals can foster the development of both indigenous suppliers (through expanding their output) and indigenous final good production (through lowering their costs for intermediate inputs).

Related models also discuss the possibility that firms that are linked through production inputs may tend to agglomerate geographically (Krugman and Venables 1995). The existence of input-output linkages and imperfect competition creates positive externalities that benefit the agglomeration of industries in particular regions. These agglomerations can occur both within narrowly defined industrial sectors only or across all industries.

The importance of interfirm linkages has also been emphasized in the literature on technology transfers from multinationals to indigenous firms (Javorcik 2004; Moran 2001). The main argument is

that in the case where firms are interlinked, local firms can improve their productivity as a result of forward or backward linkages with multinationals affiliates. For example, local firms supplying inputs to multinationals may be able to improve the quality of their output as multinationals set high quality standards for them and, as case study evidence for developing countries shows, may even instruct them in the use of advanced production technology.

Measuring interfirm linkages is, at least in theory, straightforward, as it involves determining the amount of inputs purchased from local suppliers, or output sold as intermediate inputs to domestic customer firms (e.g., Görg and Ruane 2000). A number of issues arise, however. First, these absolute values need to be expressed relative to some firm characteristic in order to be comparable across firms. For forward linkages it is generally expressed as locally sold output relative to total output, while backward linkages are usually calculated by dividing local inputs by either total inputs or total output. The simple ratio of local inputs to total outputs could be misleading as an indicator of backward linkages, however. A firm with a low ratio of local inputs to output could source all of its inputs in the local economy, since total inputs are not considered in this expression. This possibility could, for example, be a particular problem in industries that have a high content of value added but only a relatively low input content, such as, for example, software development or software production.

For forward linkages, the main difficulty is to determine what amount of output is directed to other firms as an intermediate input (i.e., the linkage) as opposed to output sold directly on the market to final consumers.

Policy Implications The purpose of measuring intersectoral linkages in the spirit of Hirschman (1958) is to determine key sectors in the economy, those sectors in which resources should be concentrated to achieve the highest growth. The interfirm linkage concept analyzes the extent of integration of firms, particularly affiliates of foreign multinationals into the host economy.

Although the empirical estimation of backward and forward linkages using the interfirm approach can be seen as an approximate indicator of the degree of a firm's integration into the host economy, this measurement in itself provides only limited information, giving no consideration to the effectiveness of these linkages. For a complete analysis more relevant to policy, there would need to be an examination of their effects, for example, on the development of local firms, indirect employment generated in domestic firms, the emergence of agglomerations, or the transfer of technology to linked local firms. This issue forms an important research agenda for academics and policymakers.

See also appropriate technology and foreign direct investment; footloose production; foreign direct investment and labor markets; foreign direct investment under monopolistic competition; foreign direct investment under oligopoly; location theory; multinational enterprises; technology spillovers

FURTHER READING

Görg, Holger, and Frances Ruane. 2000. "An Analysis of Backward Linkages in the Irish Electronics Sector." *Economic and Social Review* 31: 215 35. An empirical analysis of linkages in domestic firms and multinationals using firm level data for Ireland.

Hirschman, Albert O. 1958. *The Strategy of Economic Development*. New Haven: Yale University Press. An early contribution setting out the benefits of intersectoral linkages for economic development.

Javorcik, Beata Smarzynska. 2004. "Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages." *American Economic Review* 94: 605 27. An empirical analysis of the importance of linkages between multinationals and domestic firms for technology transfer.

Krugman, Paul R., and Anthony J. Venables. 1995. "Globalization and the Inequality of Nations." *Quarterly Journal of Economics* 110: 857 80. A theoretical model showing the importance of linkages for agglomeration of production.

Lall, Sanjaya. 1980. "Vertical Inter Firm Linkages in LDCs: An Empirical Study." *Oxford Bulletin of Economics and Statistics* 42: 203 26. An early empirical analysis of linkages by multinational companies in developing countries.

Markusen, James R., and Anthony J. Venables. 1999. "Foreign Direct Investment as a Catalyst for Industrial Development." *European Economic Review* 43: 335 56. Theoretical model highlighting the importance of link ages between multinationals and domestic firms for development of local industry.

Marshall, Alfred. 1920. *Principles of Economics*. 8th ed. London: Macmillan. Discusses, among other things, the potential benefits of input output linkages between firms for development.

Moran, Theodore H. 2001. *Parental Supervision: The New Paradigm for Foreign Direct Investment and De velopment*. Washington, DC: Institute for Interna tional Economics. Presents many case studies on back ward linkages between multinationals and domestic suppliers.

Yotopoulos, Pan A., and Jeffrey B. Nugent. 1976. "In De fence of a Test of the Linkage Hypothesis." *Quarterly Journal of Economics* 90: 334 43. Early contribution to the measurement of intersectoral linkages using input output tables. There are other papers on this topic in the same issue of the journal.

HOLGER GÖRG

■ liquidity trap, the

The liquidity trap refers to a state in which the nominal interest rate is close or equal to zero and the monetary authority is unable to stimulate the econ omy with monetary policy. In such a situation, be cause the opportunity cost of holding money is zero, even if the monetary authority increases money supply to stimulate the economy, people hoard money. Consequently, excess funds may not be converted into new investment. A liquidity trap is usually caused by, and in turn perpetuates, deflation. When deflation is persistent and combined with an extremely low nominal interest rate, it creates a vi cious cycle of output stagnation and further expec tations of deflation that lead to a higher real interest rate. Two prominent examples of liquidity traps in history are the Great Depression in the United States during the 1930s and the long economic slump in Japan during the late 1990s.

Conventional Monetary Policy Ineffectiveness

An economy's monetary authority typically tries to manipulate money supply through open market operations that affect the monetary base for ex ample, buying or selling government bonds. As long as banks are legally required to maintain a certain level of reserves, either as vault cash or on deposit with the central bank, a one-unit change in the monetary base leads to more than one-unit change in money supply the ratio between the two is referred as the money multiplier and is usually greater than one. The reason for this relationship is that banks do not have any incentives to hold reserves, which typically do not earn interest, beyond the legal requirement, and therefore they lend out any excess reserves. Nonbank firms and individuals behave in parallel with the banks' behavior; they have no incentive to hold excess money or funds above transaction needs, so they invest it in interest-earning financial assets such as bonds and bank deposits. Thus excess money or funds go back to the hands of banks, leading to rounds of lending known as money creation. How ever, the important assumption for such behavior is that the nominal interest rate is positive, or not ex tremely low. In other words, money creation arises as long as the opportunity cost of holding money is greater than zero.

When the nominal interest rate is very close or equal to zero, the opportunity cost of holding money becomes zero, and economic agents banks, firms, or individuals tend to hoard money even if they have more money than they need for transaction purposes. More important, traditional monetary policy becomes ineffective in stimulating the econ omy because the money creation process does not function as theory predicts. Even when the monetary authority increases the monetary base, money supply becomes unresponsive or even falls. In such a situa tion, because the nominal interest rate cannot be neg ative, there is nothing the monetary authority can do.

When an economy falls in a liquidity trap and stays in recession for some time, deflation can result. If deflation becomes severe and persistent, people effectively expect negative inflation going forward. Accordingly, the real interest rate (which is defined as

nominal interest rate minus expected inflation) will be expected to rise. This in turn harms private investment through increased real cost of borrowing and thereby widens the output gap. Thus the economy falls into a vicious cycle. A persistent recession causes deflation, which raises real interest rates and lowers output even further, while monetary policy is ineffective.

In the case of the Great Depression in the United States, between 1929 and 1933, the average inflation rate was 6.7 percent. It was not until 1943 that the price level went back to that of 1929. In the case of the more recent Japanese slump, deflation started in 1995 and continued till 2005, although the degree of deflation was not as severe. Indeed, during the deflationary period, the average inflation rate was 0.2 percent.

Such a spiral deflationary situation is highly likely to involve failures in the financial system. Financial failure can intensify a liquidity trap because unexpected deflation increases the real value of the debt. Borrowers' ability to repay their debt, which is already weakened by the overall slump in consumption and investment, declines and banks' portfolios become burdened with nonperforming loans loans that are not repaid. Both the Great Depression and the Japanese 1990s slump involved banking failures. In such circumstances, banks often try to reduce the amount of new loans and terminate existing loans credit contraction called credit crunch in order to improve their capital conditions, which are worsened by writing off nonperforming loans. A credit crunch can feed the vicious cycle by making less capital available to potential borrowers, and therefore contracting investment and output. An increase in the amount of nonperforming loans in the overall economy can result in banks with good capital conditions becoming more even more cautious in their extensions of credit. Furthermore, in an economy with a fragile financial system, a liquidity trap can occur when the nominal interest rate does not reach zero because holding nonmoney financial assets may involve the risk of losing the assets and once the risk is incorporated, an extremely low level of the nominal interest rate would be essentially the same as zero.

Overcoming a Liquidity Trap Because conventional monetary policy becomes ineffective in a liquidity trap, other policy measures are suggested as a remedy to get the economy out of the trap. The monetarist view suggests quantitative easing as a solution to the liquidity trap. Quantitative easing usually means that the central bank sets up a goal of high rates of increase in the monetary base or money supply and provides liquidity in the economy so as to achieve the goal. It has been argued, for instance, that the Great Depression was caused and aggravated by the misguided policies of the Federal Reserve Board, that is, monetary contraction subsequent to the stock market crash in 1927 (Friedman and Schwartz 1963). According to this viewpoint, unconventional money easing or money gift, in Friedman's words would be the appropriate policy measure. Between 1933 and 1941, the U.S. monetary stock increased by 140 percent, mainly through expansion in the monetary base. More recently, after lowering the policy target rate to zero in February 1999, the Bank of Japan implemented quantitative easing policy and set a goal for the reserves available to commercial banks from March 2001 through March 2006.

The monetarists also suggest other unconventional market operations that include the direct purchasing by the monetary authority of other financial assets such as corporate papers and long-term foreign and domestic bonds. They argue that purchasing merely short-term assets in open market operations does not function as a remedy to a liquidity trap. The idea is that because long-term bonds and securities are still assumed to be imperfectly substitutable to short-term assets even in a liquidity trap situation, the former can be purchased in open market operations to drive the long-term interest rate down.

In the Keynesian view, expansionary fiscal policy is the conventional measure to overcome a liquidity trap; the government can implement deficit spending policy to jumpstart the demand. A typical example of expansionary fiscal policy is the implementation of the New Deal policy by President Franklin Roosevelt in 1933. This policy included

public works programs for the unemployed, including the Tennessee Valley Authority project. In the case of the Japanese liquidity trap, the Japanese government spent about ¥100 trillion (equivalent to 20 percent of GDP in 2005) for a series of public works programs over the course of a decade.

See also Bank of Japan; banking crisis; debt deflation; Federal Reserve Board; money supply; Mundell-Fleming model; quantity theory of money; seigniorage

FURTHER READING

Blanchard, Olivier. 2006. *Macroeconomics*. 4th ed. New York: Pearson/Prentice Hall. One of the chapters of this book presents a basic IS LM framework to explain the mechanism of a liquidity trap and analyzes the U.S. Great Depression and the Japanese 1990s recession.

Friedman, Milton, and Anna J. Schwartz. 1963. *A Monetary History of the United States*. Princeton, NJ: Princeton University Press. This book is a seminal work in which the authors examine the development of the U.S. economy in the post Civil War era. The biggest con tribution of this book is their monetarist views on the cause of the U.S. Great Depression; they argue that the Federal Reserve Board is responsible for worsening the Great Depression by implementing monetary contrac tion immediately after the stock market crash in New York.

Krugman, Paul. 1998. "It's Baaack: Japan's Slump and the Return of the Liquidity Trap." *Brookings Papers on Economic Activity* 2: 137 87.

———. 2000. "Thinking about the Liquidity Trap." *Journal of the Japanese and International Economies* 14 (4) (December): 221 37. In these two papers, Krugman presents his Keynesian views on the Japanese recession of the 1990s and policy solutions to it. The author em phasizes the importance of creating expected inflation to get the economy out of the liquidity trap situation and doubts the effectiveness of any monetarist policies such as quantitative easing.

Svensson, Lars E.O. 2001. "The Zero Bound in an Open Economy: A Foolproof Way of Escaping from a Li quidity Trap." *Monetary and Economic Studies* 19 (S 1): 277 312. The author agrees with Krugman in that creating expected inflation will help the Japanese econ omy out of a liquidity trap. However, he clearly differs

from Krugman's view in that he proposes a fixed ex change rate policy to create expected inflation until the liquidity trap situation disappears.

HIRO ITO

■ location theory

Location theory addresses the important questions of who produces what goods or services in which locations, and why. As many government policies involve attempts to shift production, one must first examine the basis for the initial location deci sions in order to understand the impact of altering incentives.

Nearly 200 years ago, the primary concern of early location theorists, most notably Johann-Heinrich von Thünen (1783 1850), was the optimal location of cities and farms, balancing both land costs and transport costs. In von Thünen's model, concentric rings of agricultural activity develop around a city. The production of perishable goods and/or goods needing to get to market quickly locate in the rings closer to the city, and other activities such as ranch ing locate in outer rings. Since von Thünen, many other scholars have proposed more complex location models, incorporating the production of industrial and agricultural goods and services.

Many of the questions addressed in location the ory are highly relevant to international economics. For example, trade theory explains patterns of in ternational production and trade. Similarly, much of the research on foreign direct investment (FDI) looks at where multinational firms locate various activities. Policy applications of location theory have examined ways in which different countries, states, and regions can actively compete to be production locations for both trade and FDI. Before turning to the applica tions of location theory in international economics, it is important to review the basic elements of location theory.

Standard Assumptions Most theories of eco nomic location start with these main assumptions:

1. The production process for particular goods is uniform, independent of locations.

Producing corn requires a certain amount of a particular quality of land, farm machinery, chemicals, climate, etc. Therefore, some locations are more suitable for producing corn than others. Factors of production cannot be substituted for one another. For example, superior farm machines cannot substitute for scarce land to grow corn in a big city.

2. The demand for products is separated from the production, or supply, of the products. Corn producers want to put the money they earn from farming into banks in cities. Bankers in cities want to consume agricultural goods. Therefore, transportation costs affect where goods are produced. If both Iowa and Nebraska have the same amount of corn-producing factors of production, but most of the demand for corn is in New York City, other things being equal, Iowa's shorter distance to New York makes Iowa the better location for producing corn.

3. Factors of production are immobile. Inhabitants of New York City cannot import Iowa's cheap land to grow corn locally, and Iowans cannot import New York's bankers. While some factors (capital, migrant workers) are in fact mobile, land and most natural resources are not.

Theories based on these assumptions generate the clear prediction that, to minimize production and transportation costs, certain locations will specialize in the production of particular goods and services and "export" these goods to other locations.

Trade Theory There is considerable overlap between location theory and trade theory. Krugman (1993) compares and contrasts location theory and trade theory. Differences aside, the theories are ultimately quite similar in the questions they address and the assumptions they make. Neither trade theory nor location theory is inherently international. Optimal production locations could be within one country, such as Iowa and New York City in the example already given.

Neither location theory nor trade theory identifies the specific countries or regions in which production

of particular goods will be located. Trade theory speaks to characteristics of production locations, such as relative endowments of factors of production required to produce particular goods, or comparative advantage in producing one good relative to another. Location theory speaks to the optimal location of production given the cost of factors of production and transportation costs to consumers. In order to determine specifically which countries or regions will specialize in producing which goods, researchers have tested these theories using data on the production characteristics of particular goods, the factor endowments of countries, and transportation costs. In *Sources of International Comparative Advantage*, Leamer (1984) tests the precise relationships specified in neoclassical trade theory and predicts the specific countries in which production for export should be located. The evidence is, at best, an imperfect fit with the predictions of both location theory and neoclassical trade theory.

Location of FDI Location theory is frequently applied by researchers wanting to understand factors that influence where multinational firms (MNCs) choose to locate their foreign operations. Typically lacking data on the production costs of individual MNCs or data on an MNC's costs in countries where the firm does not operate, researchers tend to study features of the different locations that should make them more or less attractive, such as a country's corporate tax rate. These features include factors related to a country's policy, economy, and technology environments, and strategic considerations for individual MNCs. These are discussed in detail below.

Host-Country Policies The policy environment of a country has an important impact on its attractiveness. Policies such as corporate tax rates affect the costs of an MNC. Hence disproportionate amounts of FDI (relative to gross domestic product, GDP) occur in low-tax countries such as Ireland and tax havens such as the Cayman Islands. Conversely, some countries put restrictions on FDI such as not allowing wholly owned subsidiaries, requiring technology transfer, limiting the repatriation of profits, and mandating a certain proportion of added value

be produced in the local market. Other things being equal, countries with more restrictive policies toward FDI tend to be less attractive FDI locations. Despite a global trend of liberalization of policies toward FDI (OECD 2006), restrictions continue in some form in almost every country.

Other features of a country's policy environment also affect its attractiveness. Labor market institutions such as the degree of unionization affect FDI flows. Similarly, the protection of property rights and intellectual property, strong institutions for contract enforcement and capital market governance, environmental regulation, and trade policy may also affect MNCs' FDI location choices (Javorcik and Spatareanu 2005).

Recently, researchers have studied various types of political risk as influences on MNCs' FDI location decisions. MNCs make location decisions based on expected future profits, and greater risk creates greater uncertainty with regard to future income streams. Thus, other things being equal, risk reduces FDI inflows. Countries that are politically risky in the sense of having a history of expropriating FDI, weak institutions, endemic corruption, autocratic governments, periodic coups d'état, or ethnic tension tend to receive negligible FDI flows (or far less than would otherwise be expected). Many of the poorest countries in the world, for example, Haiti, Honduras, and much of sub-Saharan Africa, are beset with political risk and receive very little FDI. To a greater or lesser extent, some of the aforementioned risks are present in many countries, some of which, for example, China and Nigeria, are recipients of large FDI inflows, as other things are not equal. Nigeria is one of the most oil-rich countries in the world, and China has the largest population in the world and a rapidly growing economy. These factors offset some of the risk MNCs encounter in China and Nigeria. However, if these countries were less risky, they likely would enjoy even larger inflows of FDI.

Economic Environment Features of a country's economic environment are also important determinants of FDI location. Capital, and human and natural resource endowments such as labor in China and oil in Nigeria, continue to be significant influences on FDI location choices. The importance of neoclassical factors of production (e.g., land, labor, capital, raw materials) to MNCs' FDI location choices is consistent with the predictions of location theory.

Factors of production can be intrinsic or created. The Silicon Valley area receives large inflows of FDI due in part to the quality of its human capital. The geographically concentrated pool of specialized human capital arose, in part, from the presence of leading-edge domestic companies and the entry of foreign companies. Hence, specialized labor in Silicon Valley can be viewed as a created and dynamic factor of production: created in the sense that high-tech workers developed their knowledge through their experience working at different Silicon Valley firms, dynamic in the sense that the entry of more firms into the area expanded the skill set and enlarged the pool of specialized labor.

The size of a country's economy is another important economic influence on FDI location decisions. Market size can be viewed as an indirect measure of transportation costs, since it reflects the ability of a firm to reach many consumers at a relatively low cost. The attractiveness of large countries is particularly true when an economy is both large in terms of absolute size (GDP) and wealthy in terms of GDP per capita. Brainard (1997) finds that proximity to customers is an important factor in an MNC's FDI location choices.

Finally, like political risk discussed above, economic risk is also an important influence on FDI location decisions. Countries that have experienced very high rates of inflation or currency exchange rate shocks may have difficulty attracting FDI. Similarly, other things being equal, highly indebted countries and countries with periodic fiscal crises are typically not recipients of large FDI inflows.

Technology and Agglomeration Economies Studies of the effects of spillovers, external economies, or agglomeration economies on FDI location choices date back to Alfred Marshall (1920). Agglomeration economies are benefits that accrue to firms that locate in geographically concentrated areas or "clusters" such as Silicon Valley.

Several types of benefits can arise when firms collocate. First, a geographically concentrated cluster of activity in a particular sector creates a specialized pool of skilled labor that can lower a firm's search and training costs. Second, due to labor mobility and social networks, firms can potentially gain some knowledge about the proprietary technology and processes of their competitors. At the same time, of course, firms also risk losing some of their own proprietary knowledge in this context. Third, specialized suppliers often locate near clusters, again, lowering firms' costs and giving firms more choice in make-or-buy decisions. Fourth, when clusters exist, firms and states often make significant investments in infrastructure development such as building roads, upgrading airports, and improving local universities. These features of industrial clusters create an environment where firms can potentially reap benefits larger than their direct costs.

The existence of spillovers and their influence on the location decisions of MNCs has been a controversial topic that has received considerable attention from academic researchers. Since spillovers are nearly impossible to measure, their very existence is controversial. Firms in the same sectors often do tend to locate in clusters. Some researchers take this empirical fact as prima facie evidence for the existence of agglomeration economies. However, Head, Ries, and Swenson (1995) point out that clusters could exist for many reasons that have nothing to do with spillovers. Firms might collocate for economic reasons such as the local presence of factors of production or strategic reasons such as the ability to better monitor competitors. Local governments wanting to attract high-quality jobs might offer subsidies to firms in particular industries. Thus the fact that firms do cluster is not, in itself, evidence of spillovers.

Another controversy in the literature on agglomeration economies is the question of what, if any, effect they have on the production cost of firms and hence firms' decisions to collocate. If knowledge spillovers exist, why would a leading-edge firm risk locating close to competitors that could free-ride on its investments in proprietary technology? Similarly, clusters often do eventually result in crowding. After

experiencing rapid growth, the Silicon Valley area became such a high-cost location for land and labor that many firms wanting to move there were ultimately priced out of the market. Crowding should eventually increase the production costs of all firms, including early entrants into the cluster.

Spillovers of knowledge have a less straightforward effect on firms' costs in the sense that some firms might gain and others might lose. By focusing on micro mechanisms such as buyer-supplier relationships between firms, recent research allows us to better understand which firms would benefit from knowledge spillovers. Since firms do tend to collocate, and since proprietary technology is so important to firms, further research in this area will help us better understand the complex links between spillovers, production costs, and FDI location decisions.

Firm Strategy Strategic factors that influence the location choices of MNCs include the need to locate near important clients or customers, to locate close to key rivals in order to monitor their actions, or to deter the entry of key rivals. For example, service firms such as advertising agencies often locate their foreign operations near important clients.

This literature uncovers some fundamental motives underlying firms' FDI location decisions. First, and perhaps most important, although FDI flows are often discussed in the aggregate by economists and policymakers, it is important to remember that these aggregate flows represent the sum of a great many decisions made by individual MNCs. Since firms are so different from one another, clearly, their own state variables will be very important in determining where they choose to locate. Second, the heterogeneity of firms implies that all the aforementioned characteristics of locations will not be equally attractive to all firms. Advertising agencies probably will not care about the presence of natural resources such as coal, and even manufacturing firms might not care about labor or natural resources if they are trying to decide where to locate a research and development (R&D) laboratory.

Thus, in focusing on micro mechanisms underlying FDI location choices, recent research on firm strategy has considerable potential to shed light on

which location characteristics will be valued by what type of MNC. Moreover, by looking at different kinds of location decisions such as the location of manufacturing plants and the location of R&D, this research should contribute valuable insight into the conditions under which some location characteristics may be more or less important. The recent availability of large new data sets on multinational firms such as the data from the U.S. Department of Commerce, Bureau of Economic Analysis, should help advance research in this area.

Empirical Research on FDI Location Empirical studies of FDI location have typically used one of two general research designs. First, using aggregate data on FDI flows into different countries, researchers have studied the correlation between aggregate FDI flows and many of the location characteristics discussed earlier. The problem with this type of research design is that FDI locations are chosen by individual firms, and aggregate data do not give us much insight into the micro mechanisms underlying firms' choices.

The second type of research design uses micro data on multinational firms to examine decisions to enter particular markets. Entries are generally observed as discreet data points. For example, an MNC that has operations in Canada and the United Kingdom in 1989, and then has operations in Canada, the United Kingdom, and France in 1990, has "entered" France. Although entry data give us much more insight into decisions at the firm level, there are two problems with using this type of research design in studies of FDI location. First, with only a few exceptions, publicly available data on MNCs that contain reliable information on entries often do not contain information on the size of the entry or the purpose of the entry (e.g., production facility, R&D lab, distribution operation, etc.). Second, and perhaps more important, most location decisions made by MNCs are incremental. MNCs might choose, for example, to reduce the scope of their existing operations in Canada and double the scale of their existing facilities in India. If the researcher only observes countries in which the MNC has operations, no location change would be noted. In this example,

however, there has been a fundamental shift in the locations in which the MNC operates. In the day-to-day operations of MNCs, the decision to enter a new market is made relatively infrequently. In contrast, MNCs with established subsidiaries in many different markets frequently make decisions about how much to reinvest in their existing foreign operations.

A consequence of the widespread use of entry studies in research on FDI location is that we know very little about how the location characteristics that are correlated with MNCs' entry decisions affect decisions about growing or shrinking existing operations at the margin (see Feinberg and Keane 2001). From a policy standpoint, most countries are at least as concerned about retaining, growing, and upgrading existing FDI as they are concerned about luring new FDI. Therefore, additional research on MNCs' incremental FDI location choices could contribute significantly to FDI policy.

New Location Theory: Random Chance and Time Recently, researchers applying location theory to international trade and FDI have begun to incorporate important new theoretical developments such as the role of chance and dynamics. Random chance plays a critical role in determining real-world patterns of production. For example, Krugman (1991) traces the location of the carpet industry in Dalton, Georgia, to Catherine Evans, who in 1895, by chance, made a tufted bedspread as a wedding gift. The bedspread was regarded as so beautiful that neighbors began demanding tufted items. From that beginning, with the addition of modern technology, the carpet industry grew and became concentrated in Dalton. A general theory of location or trade could not account for the chance events that give rise to global production locations such as carpets in Dalton. Rational planning, rather than historical accident, would be more consistent with the predictions of location or trade theory. The Dalton, Georgia, carpet industry example shows how chance events combine with dynamic factors such as new or increased demand, technological and infrastructure development, and the entry of new firms to create permanent global production locations. Chance and

dynamics are both present in the Dalton example because the bedspread maker could have been born anywhere.

In contrast, agglomeration economies often do not arise by chance, but they do have a dynamic component. For example, the San Francisco Bay area location of the Silicon Valley tech cluster is much less surprising and random than the Dalton, Georgia, carpet location. The Bay area is home to several of the top research universities in the world and numerous government research laboratories. The state of California has a large and thriving defense industry. Thus both demand and supply conditions favored the creation of the Silicon Valley technology cluster.

A final new contribution to the literature on location theory, FDI, and trade is the literature on hysteresis, or path dependence. Hysteresis occurs when temporary economic shocks such as exchange rate volatility result in permanent transformations in global patterns of production and trade. For example, many Japanese auto and auto parts manufacturers that exported cars to the United States endured considerable economic hardship in the 1980s when the U.S. dollar began to decline. Since there was no certainty as to when the dollar would stop falling, many firms eventually decided to relocate production from Japan to North America. Thus the global pattern of production and trade was permanently changed as a result of transitory exchange rate shocks (see Baldwin and Krugman 1989).

The idea that temporary changes in the economic environment can cause permanent changes in global production patterns potentially gives significant discretion to policymakers to use short-term policies such as subsidies and tax breaks to attract global production. However, as we discuss below, when the use of such policies is widespread both within countries and internationally, bargaining power tends to shift from states to firms, and billions of dollars in dead-weight economic loss arise every year from competition to redistribute global production.

Policy Applications In a policy context, researchers have studied the following question: What can policymakers to do to enhance the attractiveness of their locations and thus increase the likelihood

multinational firms will decide to choose the particular location? Much of the work in this area evolved from two important areas of research.

First, New Trade Theory (see Krugman 1994) introduced economies of scale into theories of international trade. In the extreme, some industries such as airplane manufacture and semiconductor fabrication are so costly to set up that the world might support only a very small number of firms. In such a case, a country could "win" a two-country game in which both countries decide whether to subsidize its own firm so that firm is able to lower its cost and take the entire world market for producing a particular good (see Brander and Spencer 1985). This type of policy is often referred to as "strategic trade policy." Despite there being no empirical evidence for the existence of industries with such extreme production structures that support only a single global firm (suitable subsidy targets), many billions of dollars are wasted every year by countries and states in subsidy games designed to "win" a larger share of global production. Knowing countries and states are willing to offer payments to lure and keep production gives firms an incentive to threaten to move. In many cases, these threats lead to bidding wars that reduce firms' costs at the expense of countries and states.

Second, Porter's *Competitive Advantage of Nations* (1998) directly considered a positive role for governments in making locations more attractive. Unlike proponents of strategic trade policy, Porter advocated broader types of policy actions to enhance the attractiveness of a location. These actions include increased spending on education and R&D, government spending on infrastructure, institutional development, and the like. This goal of these policies would be to create Silicon Valley like clusters by improving the quality of domestic factors of production such as labor and capital. Since high-quality factors of production can be deployed across many different sectors, there is much less waste involved in policies that improve factor quality than in policies that merely seek to redistribute the location of existing economic activity.

Location theory addresses the important questions of who produces what goods or services in

which locations, and why. Ideas from location theory have been widely used in international economics, in particular, to predict which countries will specialize in the production of certain goods for export, and which countries MNCs will choose as production locations. In early location models, factors of production were given rather than created, and technologies for producing particular goods were unalterable. Clearly, modern applications of location theory to trade, FDI, and international economic policy have considerably expanded the scope of these location models. However, fundamental insights from the earliest theories of location such as the importance of factors of production, trade, and transportation costs remain powerful explanations for the location of global production today.

See also foreign direct investment: the OLI framework; New Economic Geography

FURTHER READING

Baldwin, Richard, and Paul Krugman. 1989. "Persistent Trade Effects of Large Exchange Rate Shocks." *Quarterly Journal of Economics* 104 (4): 635 54. Examines the is sue of hysteresis and trade: the notion that permanent changes in the location of production and trade can re sult from transitory economic shocks.

Brainard, S. Lael. 1997. "An Empirical Assessment of the Proximity Concentration Trade off between Multi national Sales and Trade." *American Economic Review* 87 (4): 520 44. Finds evidence that proximity to con sumers is an important consideration for the location decisions of multinational firms.

Brander, James A., and Barbara J. Spencer. 1985. "Export Subsidies and International Market Share Rivalry." *Journal of International Economics* 18 (1 2): 83 100. One of many fundamental papers that Brander and Spencer have done relating to strategic trade theory; provides a theoretical background for policy related papers on strategic trade theory.

Feinberg, Susan, and Michael Keane. 2001. "U.S. Canada Trade Liberalization and MNC Production Location." *The Review of Economics and Statistics* 83 (1): 118 32. In this and several other papers, Feinberg and Keane ex amine incremental location decisions of multinational firms, as opposed to entry decisions.

Head, Keith, John Ries, and Deborah Swenson. 1995. "Agglomeration Benefits and Location Choice: Evi dence from Japanese Manufacturing Investments in the United States." *Journal of International Econom ics* 38 (3 4): 223 47. Examines whether clustering is caused by spillovers or other unobservable country and industry characteristics such as factor endow ments.

Isard, Walter, and Merton J. Peck. 1954. "Location Theory and International and Interregional Trade Theory." *Quarterly Journal of Economics* 68 (1): 97 114. A fun damental paper on location theory.

Javorcik, Beata Smarzynska, and Mariana Spatareanu. 2005. "Do Foreign Investors Care about Labor Market Regulations?" CEPR Discussion Paper No. 4839. London: Centre for Economic Policy Research. This article, and other research by Javorcik, examines many different features of countries' economic and policy en vironments and their relation to the location of FDI.

Krugman, Paul R. 1991. *Geography and Trade*. Cambridge, MA: MIT Press. A nontechnical overview of the linkages between international trade and location theories.

———. 1993. "On the Relationship between Trade The ory and Location Theory." *Review of International Eco nomics* 1 (2): 110 22. Compares and contrasts location theory and trade theory, finding that despite some dif ferences, the theories are asking similar questions and making similar assumptions.

———. 1994. *Rethinking International Trade*. Cambridge, MA: MIT Press. An introduction to New Trade Theory, the intersection of industrial organization economics and neoclassical trade.

Leamer, Edward E. 1984. *Sources of International Com parative Advantage: Theory and Evidence*. Cambridge, MA: MIT Press. An in depth and extensive examination of how to test the predictions of neoclassical trade theory. Uses a structural estimation to test the predictions of trade theory using an extensive cross country data set containing detailed information on factor endowments, production, and trade.

Marshall, Alfred. 1920. *Principles of Economics*. London: Macmillan.

OECD. 2006. *International Investment Perspectives*. Paris: OECD. Source of information on the worldwide trend toward liberalization of FDI policies.

Porter, Michael E. 1998. *The Competitive Advantage of Nations*. New York: Free Press. Combines industrial organization, international trade, and economic geography to investigate why some countries are systematically more competitive at producing particular goods and services.

SUSAN FEINBERG

■ Louvre Accord

The Louvre Accord is an agreement reached in Paris on February 22, 1987, among the Group of Six countries (the United States, Canada, Japan, Germany, France, and the United Kingdom) to stabilize the value of the dollar. Since the Plaza Accord in September 1985, the nominal dollar exchange rate against other major currencies had fallen more than 25 percent. The goal of the Louvre Accord was to arrest the decline of the dollar and to stabilize it around the prevailing levels.

Although the U.S. dollar began falling in 1985, the U.S. current account deficit in 1986 and 1987 remained large. United States treasury secretary James Baker sought to correct the imbalances by encouraging trading partners to stimulate their economies and therefore to purchase more U.S. exports (Frankel 1994). If trading partners did not want to grow more quickly, he would allow the dollar to fall more, which would stimulate U.S. net exports while reducing the net exports of America's trading partners. Baker secured an agreement during the Tokyo Summit of May 1986 to use macroeconomic indicators such as gross domestic product (GDP) growth and inflation to guide Group of Seven policy discussions. This indicator approach was intended to prompt the Japanese and Europeans to stimulate their economies in order to reduce U.S. current account deficits.

Although the Japanese and Europeans viewed this approach as an intrusion into their domestic policy arena, the Japanese, in particular, were willing to accept it in order to obtain relief from the strong yen (*endaka*). The yen had appreciated from almost 240 to the dollar before the Plaza Accord to 155 to the

dollar a year later. Profit margins of Japanese exporting firms had fallen drastically, and Japanese officials did not want the yen to appreciate further.

U.S. policymakers were also concerned about the dollar depreciating further. The U.S. current account deficit had reached 3.5 percent of GDP and was financed heavily by foreign capital flows. United States officials were concerned that expectations of further depreciations could make dollar assets less attractive to foreign investors and force U.S. interest rates to increase. Such a destabilizing shift in portfolio preferences was sometimes referred to as the "hard landing" scenario (see Marris 1985; Frankel 1994).

Components of the Louvre Accord The Louvre Accord, as negotiated a few months later, stipulated:

1. The United States would reduce its budget deficit from 3.9 percent of GDP in fiscal year 1987 to 2.3 percent in fiscal year 1988.
2. Japan would cut its policy interest rate (discount rate) by 0.5 percent and enact a fiscal stimulus package.
3. France and Germany would cut taxes.
4. All countries would fight pressure from interest groups to enact tariffs and other impediments to free trade (Meyer et al. 2002).

The participants agreed in the accompanying statement to keep exchange rates within ranges broadly consistent with economic fundamentals (Funabashi 1989). The closing rates from February 20, 1987, were to be used as midpoints. Appreciations or depreciations of the dollar by 5 percent or more would trigger obligatory consultation on policy responses, appreciations or depreciations of 2.5 percent could lead to voluntary mutual intervention, and appreciations or depreciations between 2.5 and 5 percent would lead to progressively stronger interventions.

Despite heavy intervention by the United States and Japan, the dollar fell another 10 percent against the yen over the next two months. James Baker and Paul Volcker, then the chairman of the Federal Reserve, became increasingly concerned about a hard landing. In April 1987 Baker expressed the most alarm he ever had about the fall of the dollar. Volcker

warned Congress of the incipient dangers and emphasized the need to cut the budget deficit (Funabashi 1989).

Germany and Japan While the United States made slow progress in reducing its budget deficit, Baker lobbied hard for Japan and Germany to stimulate their economies. Japan proposed an urgent stimulus package. German officials demurred, however, claiming that their economy had already reached its potential output level. The German central bank (the Bundesbank) even raised the discount rate in September 1987. The unwillingness of Germany to stimulate its economy led to open conflict between Baker and German finance minister Gerhard Stoltenberg (see James 1996). On September 27, 1987, Baker said that surplus countries would have to pay more attention to their macroeconomic indicators, meaning that they would have to stimulate their economies more to help the United States reduce its current account deficit. On October 18, 1987, he again called on Stoltenberg to undertake stimulative policies and threatened to let the dollar depreciate further if Germany did not. When the market opened the next day, the Dow Jones industrial average fell 23 percent.

James (1996) notes that, before the confrontation between the United States and Germany, the Bank of Japan had been contemplating raising the discount rate to slow down a developing bubble. After watching the conflict unfold, however, Japan went along with U.S. demands to pursue expansionary policies. In the face of this pressure from the United States, monetary policy in Japan was perhaps too expansionary (Bernanke and Gertler 1999). In 1989 and 1990 the Bank of Japan finally raised interest rates to prick the well-developed asset price bubble. Share prices then fell 90 percent between 1989 and 1992 and remained stagnant thereafter. The bursting of the bubble was disastrous for Japan because of its bank-based financial system. Falling asset prices eroded bank capital and restricted banks' ability to extend credit (Okina and Shiratsuka 2003). This in turn contributed to the "lost decade" of the 1990s, a period during which the Japanese economy remained stagnant.

Following these events, disillusionment with international policy coordination grew. Many came to view coordination efforts as thinly disguised attempts by the United States to get other countries to rescue it from the effects of its own unwise domestic policies (McCallum 1996). Even mild forms of macroeconomic policy coordination became difficult for developed economies to achieve.

See also balance of payments; Bonn Summit; Federal Reserve Board; international policy coordination; Plaza Accord; Smithsonian Agreement

FURTHER READING

Bernanke, Ben, and Mark Gertler. 1999. "Monetary Policy and Asset Price Volatility." In *New Challenges for Monetary Policy*. Kansas City: Federal Reserve Bank of Kansas City, 77 128. A discussion of the relationship between monetary policy and asset price bubbles.

Frankel, Jeffrey. 1994. "The Making of Exchange Rate Policy in the 1980s." In *American Economic Policy in the 1980s*, edited by Martin Feldstein. Chicago: University of Chicago Press. A readable and insightful history of exchange rate policy in the 1980s.

Funabashi, Yoichi. 1989. *Managing the Dollar: From the Plaza to the Louvre.* Washington, DC: Institute for International Economics. Provides a thorough, behind the scenes account of how the G5 countries were able to agree on the Plaza Accord.

James, Harold. 1996. *International Monetary Cooperation since Bretton Woods.* Oxford: Oxford University Press. A scholarly account of international monetary affairs from 1945 until the early 1990s.

Marris, Stephen. 1985. *Deficits and the Dollar. The World Economy at Risk.* Washington, DC: Institute for International Economics. Argues that America's deficits in the 1980s were unsustainable.

McCallum, Bennett. 1996. *International Monetary Economics.* Oxford: Oxford University Press. Covers international monetary issues.

Meyer, Laurence, Brian Doyle, Joseph Gagnon, and Dale Henderson. 2002. "International Coordination of Macroeconomic Policies: Still Alive in the New Millennium?" International Finance Discussion Paper 723. Washington DC: Federal Reserve Board. Surveys models of policy coordination.

Okina, Kunio, and Shigenori Shiratsuka. 2003. "Japan's Experience with Asset Price Bubbles: Is It a Case for Inflation Targeting?" In *Asset Price Bubbles: The Implications for Monetary, Regulatory, and International Policies*, edited by William Hunter, George Kaufman, and Michael Pomerleano. Cambridge, MA: MIT Press. Examines the relationship between Japan's asset price bubble and monetary policy.

WILLEM THORBECKE

■ Maastricht Treaty

Maastricht is the city in the Netherlands where heads of government reached agreement on a document that would give birth to the European Union. The treaty represents a landmark in the process of European integration as it would define the rules of the game under which member states would eventually gain membership into the single currency area. Originally, the European Union (EU) consisted of 12 member states when the treaty first came into force in November 1993. Later, there would be further enlargements, the most recent occurring in 2004 when 10 new member states were admitted. In 2007, 2 more member states (Bulgaria and Romania) were admitted, bringing the total EU membership to 27 countries that would eventually subscribe to the treaty.

Because it was designed to deal with existing member states, not ones that would join in future, the original Maastricht Treaty could not cope with enlargement. Consequently, the Maastricht Treaty was followed by the Treaty of Amsterdam in 1997 and the Treaty of Nice in 2001. The most important change to the original Maastricht Treaty dealt with voting rules for policy approval: the Maastricht Treaty required unanimous approval of policy changes at the EU level, whereas succeeding treaties allowed for majority approval. The system in place in 2007 essentially assigned a weight to each country's influence in the EU roughly based on population size. Although some decisions require unanimity (e.g., foreign policy questions), others would henceforth require a qualified majority, that is, a vote that reflects the relative importance of some EU members based on their population (e.g., measures affecting the operations of the internal European market). Even when a qualified majority is used to make a decision, the number of votes must exceed a threshold (more than 50 percent of votes) before any decision can be adopted.

The grand objective behind the Maastricht Treaty was to put in place institutions and policies that would advance the ideal of a unified Europe. Although the centerpiece of the treaty concerned the rules intended to foster economic and monetary union, it also included social and legal provisions to harmonize existing laws and regulations among the member states and to develop common foreign and security policies. Some governments objected to particular provisions. The United Kingdom, for example, found the Social Chapter, a legal provision that dealt with policies concerning wage setting and health and safety regulations, unacceptable and opted out of the relevant provisions. Denmark rejected the treaty after putting it to a vote in a referendum. Eventually, the United Kingdom and Denmark received an opt-out agreement that would give them the option to join the monetary union in the future.

Maastricht Convergence Criteria The road map to monetary union was first laid out in the Delors Report of 1989, named after the former French politician who is also a former president of the European Commission. Critical to achieving monetary union were the so-called convergence criteria. The goal was a simple one in principle, namely to ensure that the members of the area that would form a monetary union would have comparable

macroeconomic environments. This would, first, facilitate the introduction of an unalterable fixed exchange rate, to be followed by the introduction of a common currency. The original report called for European Monetary Union (EMU) by 1999. At the end of 1995, leaders of the EU announced plans for a single currency that would be called the euro.

The road to the single currency proceeded in three stages. The first stage, which began in January 1990, was intended to liberalize the movement of capital. The second stage began in January 1994 when member states began to enact economic policies that would lead to a form of convergence as defined in the treaty. It is perhaps the so-called convergence requirements that have received the most attention among policymakers and academics over the years. These continue to form the core of the treaty. The member states that joined the enlarged EU in 2004 are also required to adhere to the convergence requirements in order to adopt the euro. Unlike Denmark or the United Kingdom, however, all new members of the EU eventually must adopt the euro.

The convergence criteria consist of five elements:

1. Inflation rates up to a maximum of 1.5 percentage points above the average of the three lowest national inflation rates.
2. Interest rates up to a maximum of 2 percentage points above the average of the three lowest national long-term interest rates.
3. Budget deficits up to a maximum of 3 percent of gross domestic product (GDP).
4. An accumulated public or government debt of up to 60 percent of GDP; and
5. Maintenance of stable exchange rates for at least two years.

Although the convergence principles seem clear, they are not unambiguous. Over the years countries have haggled over the exact definition of stability in the exchange rate and what could or could not be included in the deficit calculations. Disputes over the precise interpretation of the deficit principle eventually led to a separate agreement known as the Stability and Growth Pact, which laid down a series of fiscal rules that themselves are subject to continuing debate. Similar concerns arose with respect to the

debt to GDP ratio principle. The treaty suggests that if the 60 percent threshold is not met, the ratio should show signs of diminishing sufficiently in the future and, consequently, approach the reference value of 60 percent at a "satisfactory pace." The wiggle room in the convergence criteria allowed Belgium, Italy, and Greece to join the euro area. It is, therefore, clear that the inflation convergence requirement is first among equals in the list of Maastricht convergence principles.

The third stage began in January 1999 with the establishment of a European Central Bank and culminated in the introduction of the single currency in January 2002.

See also common currency; euro; European Central Bank; European Monetary Union; European Union; optimum currency area (OCA) theory

FURTHER READING

De Grauwe, Paul. 2005. *The Economics of Monetary Union.* 6th rev. ed. Oxford: Oxford University Press. Covers the history and the case for or against monetary integration as well as the economics of monetary unions more generally.

European Union. *Treaty of Maastricht on European Union.* Downloadable at http://europa.eu/scadplus/treaties/maastricht en.htm. The text, written in legal language, of the actual Maastricht Treaty.

Fratianni, Michele. 1992. "The Maastricht Way to EMU." Princeton Essays in International Finance No. 187. Princeton, NJ: Princeton University. Describes the essential events and issues surrounding the decision to adopt convergence requirements in the Maastricht Treaty.

PIERRE L. SIKLOS

■ market access

Acquiring market access is one of the main motives that drive multinational firms to invest abroad. To serve foreign markets, firms are often faced with two choices: exports versus foreign direct investment (FDI). In the presence of large trade costs (e.g., tariff and transportation costs), firms are less inclined to

export to the foreign market. Instead, they are more likely to supply foreign consumers through local production. This incentive to avoid trade costs by producing inside a foreign market is also referred to as the tariff-jumping motive. As a result, multinational firms duplicate their production in the foreign countries and expand horizontally. The FDI arising from this consideration has been defined by the researchers as the horizontal type of FDI. This horizontal type of FDI is distinguished from vertical FDI, in which a firm shifts production to a low-cost location and serves all markets from there.

The Proximity-Concentration Trade-off The theoretical literature on horizontal FDI is led by the seminal work of Markusen (1984), one of the first authors to show from an international trade perspective that multinationals may arise endogenously in equilibrium. A key ingredient of the models in this literature is the interaction of trade costs and plant-level scale economies. While multinational firms are able to avoid trade costs by operating local production plants in foreign countries, they must pay a plant-level fixed cost for each additional plant. In contrast to the trade costs that prompt firms to produce the same goods and services in multiple countries, the plant-level scale economy encourages firms to concentrate their production in one location. Hence, as predicted in these studies, multinational activity is more likely to arise in industries that require a relatively small plant-level fixed cost and in countries that impose large trade costs on imported goods. Moreover, multinational activity should be concentrated among countries that are similar in both market size and factor endowment.

These predictions have been formally examined in several important empirical studies, including Brainard (1997); Carr, Markusen, and Maskus (2001); and Yeaple (2003) among others. These papers formally examine the extent to which the decision by multinationals to invest overseas is affected by their motive of maximizing proximity to consumers and the trade-off in plant-level scale economy.

Multinational firms' market access motive is broadly confirmed. Based on mainly the U.S. inward or outward FDI data, these studies show that multinationals'

affiliate sales increase with both the tariff rate the foreign country imposes on the parent country and the transportation cost between the two. When the plant-level fixed cost, measured by the average size of a plant in terms of production workers, is larger, however, the affiliate sales decrease. Furthermore, Carr, Markusen, and Maskus (2001) find that the smaller the difference between two countries in market size (i.e., GDP) and factor endowment (i.e., capital labor ratio), the greater the FDI flow between the two. These results strongly suggest the existence of horizontal FDI that is motivated by market access.

Market Potential A more recent study by Head and Mayer (2004) further extends the notion of market access. They introduce the importance of market potential in multinationals' location choice and point out that, while the ability to access a foreign market at little cost motivates firms to locate production in that country, the ability to access other markets from that country also matters. In other words, the definition of "market" in the market access motive should not be limited to the market of the host country; it should include all the markets that can be more easily accessed by the host-country plant.

To test this idea, Head and Mayer (2004) construct a measure of market potential for each potential host country. This measure represents the demand of all markets weighted by the trade costs required to export to these markets from the host. Countries with better access to other markets due to, for example, geographic proximity have greater market potential. Using a sample of Japanese multinational firms, Head and Mayer (2004) offer important evidence that regions with a larger market potential indeed attract more multinationals.

The Impact of Regional Economic Integration Given the role of market access in foreign direct investment, any change in market accessibility is expected to have an impact on multinationals' activities. In particular, regional economic integration, by reducing the trade barriers between participating countries (and keeping barriers against the rest of the world intact) will likely alter the location decision of multinational firms in the integrated region.

First, a decline in trade costs between home and host countries raises the profit associated with exporting. Multinationals' market access motive may hence be weakened and even dominated by their incentive to concentrate production geographically and achieve economies of scale. As a result, FDI could be replaced by exports. For example, a great concern in Canada during the formation of the Canada-U.S. Free Trade Agreement was that U.S. multinational affiliates located in Canada would leave Canada and serve the Canadian market through exports originating in the United States. This expectation is, however, inconsistent with the evidence. The study by Feinberg and Keane (2001) offers direct empirical insights to this issue. They show that reductions in the Canadian tariff against the United States have little impact on the U.S. multinational firms' affiliate sales in Canada and, moreover, reductions in the U.S. tariff against Canada increase the U.S. multinationals' export sales from Canada to their home countries. These findings suggest that regional economic integration does not necessarily lower multinationals' incentive to invest abroad; it may in fact increase multinationals' incentive to produce in the integrated foreign country and export back to home.

Second, regional economic integration that improves the market access among member countries could also affect the location decision of outsider firms. As the benefit of improved market access is exclusive to firms located within the integrated region, outsider firms have an increased incentive to produce in a member country and supply the rest of the region through exports. The underlying logic is closely related to the notion of market potential in Head and Mayer (2004): regional economic integration increases members' market potential and thus their ability to attract multinationals. This intuitive prediction is derived and empirically examined by Ekholm, Forslid, and Markusen (2007), who emphasize multinationals' exports to third countries. As demonstrated in this study, after the formation of a preferential trading bloc both insider and outsider firms tend to adopt an export-platform FDI strategy, in which they increase their investments in some

members of the bloc and supply the other member countries through exports.

In sum, market access, or more generally market potential including the ability to better access other markets, is an important motivation for firms to engage in FDI and establish production facilities abroad. Regional economic integration, by lowering internal trade barriers, can make a region more attractive to multinational firms and lead to an expansion in export-platform FDI into the region.

See also foreign direct investment: the OLI framework

FURTHER READING

Brainard, S. Lael. 1997. "An Empirical Assessment of Proximity Concentration Trade off between Multi national Sales and Trade." *American Economic Review* 87 (4): 520 44. Examines the role of a trade off between achieving proximity to customers and achieving scale economies in multinationals' investment activities abroad and finds that overseas production by multina tionals increases when transport costs and trade barriers are higher and when investment barriers and plant level scale economies are lower.

Carr, David, James Markusen, and Keith Maskus. 2001. "Estimating the Knowledge Capital Model of the Multinational Enterprise." *American Economic Review* 91 (3): 691 708. Offers a direct test of the predictions generated in a knowledge capital model, estimates the volume of foreign affiliate production as a function of host and home country characteristics, and finds evi dence of both horizontal and vertical types of foreign direct investment.

Ekholm, Karolina, Rikard Forslid, and James Markusen. 2007. "Export Platform Foreign Direct Investment." *Journal of European Economic Association* 5 (4): 776 95. Examines theoretically and empirically the determinants of export platform FDI, in particular the effect of a free trade area between a high income and a low income country on firms' location decision, and shows that this free trade area can lead both the insider and outsider country firms to adopt an export platform FDI strategy.

Feinberg, Susan, and Michael Keane. 2001. "U.S. Canada Trade Liberalization and MNC Production Location." *Review of Economics and Statistics* 83 (1): 118 32. Ex amines how the bilateral trade flows of U.S. multina

tional corporations and their Canadian affiliates re sponded to U.S. Canadian tariff reductions and finds that the bilateral trade liberalization motivated U.S. multinationals to integrate their North American pro duction and increase their exports from Canadian affil iates to the United States.

Head, Keith, and Thierry Mayer. 2004. "Market Potential and the Location of Japanese Investment in the Euro pean Union." *Review of Economics and Statistics* 86 (4): 959 72. Examines the effect of market potential on Japanese multinational firms' location decision and finds that not only the size of the domestic market but also the size of export markets matter significantly in multina tionals' location choice.

Markusen, James. 1984. "Multinationals, Multiplant Eco nomics, and the Gains from Trade." *Journal of Inter national Economics* 16 (3 4): 205 26. Develops a gen eral equilibrium model of a multinational firm in which the productivity of a joint input is independent of the number of plants operated by a firm and finds that the existence of multinational firms increases technical effi ciency as it eliminates the duplication of the joint input among independent national firms.

Yeaple, Stephen. 2003. "The Role of Skill Endowments in the Structure of U.S. Outward Foreign Direct Invest ment." *Review of Economics and Statistics* 85 (3): 726 34. Investigates empirically the industry and country deter minants of U.S. outward foreign direct investment and finds evidence of both market access and comparative advantage motives in U.S. multinationals' investment activities.

MAGGIE XIAOYANG CHEN

■ **market efficiency**

See interest parity conditions; forward premium puzzle

■ **market size and foreign direct investment**

Market size is an important factor in determining bilateral activities of multinational firms, regardless of whether these activities are measured as foreign affiliate sales, affiliates' employment, or foreign direct investment (FDI). FDI reflects the objective of es tablishing a lasting interest by a resident enterprise in one economy in an enterprise that is resident in another economy. Discussions often focus on the market size of the host country, although the size of neighboring countries and other countries more easily reached from the host country also play a role. Market size refers mainly to the size of markets for goods and services. Although market size of the host country is widely considered, the size of the home country is also important for bilateral FDI. Larger and more advanced countries are home to more multinational firms. The pool of knowledge neces sary to generate the specific advantages that enable firms to survive in a foreign environment increases along with a country's size and level of development.

Empirical Evidence and the Gravity Equation Most FDI flows toward large markets. Since the 1980s, the United States has received the largest in flows and holds the largest FDI stocks. Among the developing countries, China has gained the domi nant position. The size effect of the host country has been the object of empirical studies since the early 1960s. The market size hypothesis states that FDI flows are positively related to the host country's market size. Various studies have found that market size and market growth affect FDI flows positively.

Gravity equations explain activities between two countries by the size of the source and the destination country and the distance between the two. Originally developed to explain international trade, the gravity equation successfully explains aggregated activities of multinational firms such as foreign affiliate sales or FDI as well. Using bilateral data at the country or sector level, about 60 percent of the variation in multinational firms' activities are explained by these three variables. The effect of the size variables (mostly gross domestic product or population) is found to be robust and positive (Brainard 1997; Buch et al. 2005). Moreover, the relationship is robust over different time periods and against changes in the analyzed sample. In particular, no systematic dif ferences are found with respect to the size variables in studies explaining FDI in developing countries

relative to FDI in members of the Organisation for Economic Co-operation and Development (OECD).

Market size as a determinant of FDI has received increasing attention because the expansion of FDI activities after 1985 has predominantly been among OECD countries. The bulk of the FDI flows involves the same countries as home and as host countries even within narrowly defined industries. Thus factor cost differentials could not have been the source of this increase in FDI activities. The motive behind the increased FDI activities has not been found in lower production cost but in improved market access for products. Thus the new theoretical explanation for the emergence of multinational firms placed goods market access and therefore market size in the center of the analysis.

FDI for Market Access According to this line of the literature, named proximity-concentration theory, or theory of the horizontal firm, the choice of a particular mode of market access drives the internationalization of firms. Firms invest abroad to serve the foreign market by production abroad instead of exporting there. Producing abroad saves distance costs and improves market access. Distance costs comprise all costs that come with segmented markets, including transportation costs, tariffs, quotas, information costs, and registration costs.

At a given level of distance costs, a firm is more likely to decide in favor of production abroad the larger its sales are in the foreign country, because distance costs savings through production abroad are per-unit savings. Sales of the firm in the foreign country in turn depend on the foreign country's market size. Thus a firm is likely to realize larger savings the larger the foreign country's market. The additional costs associated with producing in the foreign country, in contrast, are fixed costs, that is, costs independent of sales. These fixed costs can be huge, so it does not pay to replace exports by production abroad if per-unit distance cost saving are not large enough. Hence a larger country attracts more foreign multinational firms, because market size increases the per-unit cost saving but leaves the fixed costs unaffected.

While market access is important, it is certainly not the only factor determining the decision of a firm to go abroad. Factor costs are also important (Braconier, Norback, and Urban 2005). The high explanatory power of market size in the gravity equation does not necessarily mean that goods market access is the most important motive of engagement in FDI. The size of host country's *factor* markets can also affect FDI positively. A large pool of well-trained labor with specific skills drives FDI in areas such as finance or software development to London and Bangalore, respectively. The large supply of cheap labor attracts FDI in manufacturing to China.

Factor market potential also stands behind the positive effect of the *home* country's size on FDI. The relevant factor here is knowledge capital, as Markusen (1995) calls it. Knowledge capital can be proprietary product or process know-how, efficient organizational structures, a brand name, or reputation. Knowledge capital enables firms to generate the firm-specific advantage that is necessary to produce successfully in a foreign environment. Knowledge capital increases with the stage of development and with population, since knowledge is embodied in people. Hence knowledge capital, that is, the ability to generate the firm-specific advantages, rises with the market size of the home country.

Policies to Attract FDI Given that market size is such an important determinant of FDI, restrictions on market access have been used to attract FDI. That has mainly been done by excluding or obstructing other channels to supply the market, mainly trade. FDI carried out to circumvent trade barriers has been named tariff-jumping FDI. Such tariff-jumping FDI has been found to be important for Japanese FDI to the United States and to Europe (Barrel and Pain 1999). Tariff jumping has also been a motive for FDI in Latin American countries, particularly Brazil, during their period of import substitution.

Although restricting trade is an option to attract FDI, particularly for large countries, the opposite strategy has proven more successful. By regional integration, that is, through the reduction of trade barriers, countries in the European Union (EU) and in East Asia have successfully enlarged their markets

in order to attract FDI. Regional integration has increased the market potential of individual countries tremendously. Ireland, for example, is chosen as host country by many multinational firms from outside the EU that aim at supplying goods and services in the whole EU. In addition to the increase of FDI from outside, regional economic integration has increased FDI within the region. That can be observed in the EU, as well as in the North American Free Trade Area (NAFTA) and in East Asia. Reducing internal barriers to trade alters location decisions within the regions. Firms can exploit cost differences within the region, which motivates increased FDI from within the region.

In sum, there is strong empirical support for the importance of market size as a determinant of multinational activity. This has spurred theoretical research that has found market access to be a strong motive for FDI. The market access motive has received support by intensified FDI flows following market size growth through establishing regional integration areas such as the EU and NAFTA.

See also factor endowments and foreign direct investment; gravity models; location theory

FURTHER READING

Barba Navaretti, Giorgio, and Anthony J. Venables, eds. 2005. *Multinational Firms in the World Economy.* Princeton, NJ: Princeton University Press. A comprehensive overview of today's stand of economic research on multinational firms.

Barrel, Ray, and Nigel Pain. 1999. "Trade Restraint and Japanese Direct Investment Flows." *European Economic Review* 43 (1): 29 45. An empirical examination of the effect of trade protection measures on FDI.

Braconier, Henrik, Pehr Johan Norback, and Dieter Urban. 2005. "Multinational Enterprise and Wage Costs: Vertical FDI Revisited." *Journal of International Economics* 67: 446 70. An important empirical analysis of determinants of FDI based on a data set that includes many developing countries and almost all OECD countries.

Brainard, S. Lael. 1997. "An Empirical Assessment of the Proximity Concentration Trade off between Multinational Sales and Trade." *American Economic Review* 87 (4): 520 44. The first empirical assessment of the proximity concentration theory.

Buch, Claudia M., Jörn Kleinert, Alexander Lipponer, and Farid Toubal. 2005. "Determinants and Effects of Foreign Direct Investment: Evidence from German Firm Level Data." *Economic Policy* 20 (41): 51 110. An early study of microlevel evidence on multinational firms.

Markusen, James R. 1995. "The Boundaries of Multinational Enterprises and the Theory of International Trade." *Journal of Economic Perspectives* 9 (2): 169 89. The first formulation of a theory of the multinational firm that includes horizontal and vertical firms based on the idea of knowledge capital as a defining element.

. 2002. *Multinational Firms and the Theory of International Trade*. Cambridge, MA: MIT Press. A comprehensive summary of microeconomic general equilibrium theory on multinational firms.

JÖRN KLEINERT

■ Marshall-Lerner condition

The Marshall-Lerner condition establishes the circumstances in which a country's trade balance, or current account balance of payments, would improve as a result of a currency devaluation or depreciation. It was named after the economists Alfred Marshall and Abba Lerner, who are thought to have first discovered it. The theoretical basis of the Marshall-Lerner condition is the "imperfect substitutes model," which is a workhorse model of international trade that assumes that neither imports nor exports are perfect substitutes for domestic goods. It is also closely linked to the so-called elasticities approach to the balance of payments and the J-curve effect, since the answer depends on how responsive (or "elastic") imports and exports are to changes in their relative prices.

The Marshall-Lerner condition is the specific condition under which a real currency depreciation, interpreted as a fall in the relative price of exports in terms of imports, will improve the current account. More specifically, if the current account balance is initially zero neither a deficit nor a surplus the

Marshall-Lerner condition requires that the sum of the absolute values (ignoring signs) of the price elasticity of export demand, which measures the response of exports to the fall in the relative price of exports, and the price elasticity of import demand (the numerical response of import demand to the rise in the relative price of imports) exceed one. The export and import demand elasticities will be "elastic" if a 10 percent change in their prices brought about by the fall in the currency leads to a change of more than 10 percent in the quantities demanded. Moreover, the greater the amount by which the sum of these elasticities exceeds one the greater the improvement in the current account. On the other hand, a currency appreciation or revaluation will worsen the current account.

Intuitively, the reasoning is that when foreign goods become more expensive in the home country, fewer imports will be demanded and when the home country's goods become cheaper abroad, foreigners will buy more of its exports. Thus, provided the responses of buyers of imports and exports to the price changes are strong enough, the Marshall-Lerner condition will be satisfied. Export values will rise and import values will fall so the net effect will be an improvement in the current account.

Note that this is the simplified version of the Marshall-Lerner condition, which focuses on the responses of demand for exports and imports and assumes that responses on the supply side, measured by the elasticities of import and export supply, are both set to infinity. This implies that the home country is small in terms of its influence on international prices, meaning that it can buy all the imports it wishes at a given international price and is willing to sell to foreign countries as much as they want at a fixed price. The outcome is more complex if this assumption is relaxed and/or the current account is not initially zero. In addition, the predicted outcome for the current account is conditional on holding everything else constant, such as the effects of the fall in the value of the currency on incomes in both the home country and abroad.

Importance of the Marshall-Lerner Condition
The Marshall-Lerner condition is important because it enables policymakers to predict the effects of

changes in the exchange rate on the balance of payments (equivalent to the current account if capital flows are ignored). This is especially relevant to developing countries that find it difficult to sell their goods abroad because their currencies are "overvalued," or overpriced, or that wish to devalue their currencies in an attempt to reduce their current account deficits. In 1994, for example, a number of countries belonging to the African Financial Community franc zone, an arrangement under which several former French colonies fix their currencies to the French franc, devalued simultaneously by 50 percent to improve their balance of payments. But the condition is also of interest to developed countries, including the debate over the sustainability of the large U.S. current account deficit and the extent to which this can be reduced by a depreciation of the U.S. dollar.

The Marshall-Lerner condition also has important implications for the stability of the foreign exchange market. A foreign exchange market is stable if a disturbance that upsets the balance between the supply and demand for foreign exchange is corrected automatically through adjustment of the exchange rate. If, however, market movements in the exchange rate result in a movement further away from a supply and demand equilibrium, then the foreign exchange market is unstable. Although it is possible to establish conditions for the stability of the foreign exchange market based on the supply and demand curves for foreign exchange, these cannot be measured empirically. It turns out, however, that if the Marshall-Lerner condition is fulfilled, the foreign exchange market will also be stable, so all one has to do is to measure the price elasticities of demand for exports and imports.

Empirical Studies Not surprisingly, there has been a substantial amount of empirical work testing the Marshall-Lerner condition both for one country's trade with the rest of the world and for trade flows between two countries. Early work following Marshall's (1923) original publication tended to support the condition giving rise to the expression *elasticity optimism* but later studies in the 1940s and after World War II were more "pessimistic"

and suggested that the sum of the demand elasticities might be below or close to one. In the 1990s and early 21st century many of the earlier methodological and measurement problems associated with the testing of the Marshall-Lerner condition have been addressed and a consensus picture has emerged.

Relative prices do seem to play a significant role in determining exports and imports so the Marshall-Lerner condition is likely to be satisfied for most countries, but there are some important qualifications. First, although studies measuring export and import demand elasticities for samples of countries are generally supportive of the condition, other studies looking at individual countries over a long period of time have found little evidence of a systematic relationship between the trade balance and the relative prices of exports and imports. One also has to distinguish carefully between developed and developing countries and countries that have a high proportion of natural resources, including crude petroleum, and services in their exports and imports, since these structural differences in their economies may result in differences in their responses to export and import price changes. Third, it is crucial to allow for the fact that export and import demand may adjust only slowly over time to changes in relative prices. In particular, elasticities measured over a six-month period following the currency depreciation may be substantially less than one and may give rise to the "J-curve effect." Even short-run elasticities based on a one-year delay could still be quite low. Only when a longer period has elapsed can one be confident that the responses are large enough to satisfy the Marshall-Lerner condition.

Finally, one has to be careful to allow for episodes in history that may make the measured elasticities unreliable or unrepresentative of normal behavior, such as following the reunification of Germany in 1989, or the establishment of the North American Free Trade Agreement among Canada, Mexico, and the United States in 1994.

See also balance of payments; fear of floating; foreign exchange intervention; international reserves; J-curve effect; North American Free Trade Agreement (NAFTA); real exchange rate

FURTHER READING

Hooper, P., K. Johnson, and J. Marquez. 2000. "Trade Elasticities for the G 7 Countries." *Princeton Studies in International Economics* 87 (August): 1 29. An influential empirical study that is supportive of the Marshall Lerner condition and looks at the evidence for the so called Group of Seven advanced industrialized countries.

Lerner, A. 1944. *The Economics of Control.* London: Macmillan. Provides the origins of the Marshall Lerner condition.

Marshall, A. 1923. *Money, Credit, and Commerce.* London: Macmillan. Provides the origins of the Marshall Lerner condition.

Reinhart, C. 1995. "Devaluation, Relative Prices, and International Trade." *International Monetary Fund Staff Papers* 42 (2): 290 312. An influential empirical study that is supportive of the Marshall Lerner condition and focuses on the adjustment of developing countries facing large imbalances in their balance of payments.

Rose, A. 1991. "The Role of Exchange Rates in a Popular Model of International Trade." *Journal of International Economics* 30: 301 16. Provides an explanation of the relationship between the Marshall Lerner condition, the imperfect substitutes model, and the elasticities approach and an example of the empirical literature.

PETER WILSON

■ **mercantilism**

A school of economic thought developed in 16th- and 17th-century England, mercantilism argued that a country's primary economic objective should be the achievement of a trade surplus with the associated inflow of gold. The central idea was that trade was a zero-sum game and that a country could amass gold through a balance-of-trade surplus only at the expense of another country. Mercantilism is therefore a form of economic nationalism, with foreign trade used to enhance the wealth and power of one country at the expense of others.

Mercantilism is not a unified body of thought; however, Thomas Mun's *England's Treasure by Foreign Trade*, published in 1664, is typical of mercantilist writings. The policy implications of the mercantilist position included the limitation of imports and the promotion of exports. Mercantilism formed the basis for the organization of Britain's trade with its colonies, with the colonies restricted to supplying the raw materials for Britain's emerging manufacturing sector. The colonies were prohibited from competing with these so-called enumerated goods, and the entire colonial trading system was organized to promote the economic strength of the mother country and produce a trade surplus. The popularity of mercantilist thinking has been attributed to the realities of 16th- and 17th-century statecraft, in which the ability to fight foreign wars depended on having the gold resources to finance such adventures; in this case, gold accumulation through export surpluses did contribute to national power.

Adam Smith and Mercantilism The protectionist implications of mercantilism led Adam Smith to identify mercantilism as one of the erroneous systems of political economy that he sought to displace with his *An Inquiry into the Nature and Causes of the Wealth of Nations* (1776). For Smith, the mercantilists were guilty of a simple confusion about the nature of wealth, believing that it consisted of gold rather than the human and physical stock of capital. Smith believed free trade was the appropriate policy. David Hume's essay "Of the Balance of Trade" (1752) demonstrated that the objective of a balance-of-trade surplus was self-defeating, as the specie-flow mechanism would lead to domestic price level changes, which would bring the trade balance back into equilibrium. In arguing this, Hume developed an early version of the quantity theory of money. By the mid-18th century, therefore, mercantilism had been discredited and was entirely eclipsed by the theory of comparative advantage and the associated benefits of free trade developed by David Ricardo in *Principles of Political Economy and Taxation* (1821). Britain abandoned mercantilism and its associated trade restrictions in the 1840s.

Mercantilism was gradually replaced by free trade and comparative advantage as the new economic orthodoxy at the theoretical level, although one of the mercantilists' central policies protectionism continued to have its adherents, especially in Germany and the United States, the late industrializers of the 19th century. Both Friedrich List and Alexander Hamilton, influential German and American thinkers, respectively, advocated protectionist policies to spur their countries' industrialization. Even where free trade had become the orthodoxy, such as in Britain, protectionist policies returned in the face of the economic disaster of the 1930s. Even John Maynard Keynes, in *The General Theory of Employment, Interest, and Money* (1936), included the mercantilists among his predecessors in believing that an increase in the money supply could increase domestic production. Keynes argued that there was an "element of scientific proof in mercantilist doctrine" (335).

Neomercantilism In the post-1945 period, the successful late, late industrializers of East Asia, especially Japan, South Korea, and Taiwan and in the 1990s and early 21st century, China have been described as "neomercantilist." These economies ran trade surpluses during much of their post-1950 rapid growth periods, based on import restrictions, the promotion of exports, and an exchange rate pegged to the U.S. dollar (at artificially low levels, according to some). The trade surpluses enabled the East Asian economies to achieve rapid economic growth over a long period without being subject to a balance of payments constraint a constraint that affected many other developing countries' growth prospects, especially in Latin America. The prefix *neo* in *neomercantilism* therefore referred to the new period in which countries pursued trade surpluses as well as the economic, as opposed to military, definition of "national power." The term is usually used in a normative way to indicate that East Asian countries, and occasionally others, have been following nationalist policies at the expense of other countries.

See also balance of payments; beggar-thy-neighbor policies; Bretton Woods system; comparative advantage;

gold standard, international; international reserves; money supply; quantity theory of money

FURTHER READING

Heckscher, Eli. 1955. *Mercantilism*. 2 vols. London: Allen and Unwin. A classic text on mercantilist writings by one of the pioneers of modern trade theory.

Keynes, John Maynard. 1936. *The General Theory of Employment, Interest, and Money*. London: Macmillan.

Nester, William. 1990. "The Development of Japan, Taiwan, and South Korea: Ends and Means, Free Trade, Dependency, or Neomercantilism?" *Journal of Developing Societies* 6: 203 18. Examines whether the East Asian countries can be accurately described as pursuing "neomercantilist" policies.

Viner, Jacob. 1937. *Studies in the Theory of International Trade*. New York: Harper and Brothers. A historical examination of theories of international trade that includes a critique of mercantilist thinking.

PAUL BOWLES

◼ Mercosur

Traditional rivals in the southern cone of Latin America, Argentina and Brazil first explored economic collaboration in the early 1980s. For the first time in centuries, the moment was right. Both countries underwent political transitions in the mid-1980s from military rule to civilian regimes Argentina in 1983 and Brazil in 1985. The reestablishment of democratic governments provided an opportunity to seek policies that would consolidate the democratic process and limit the sphere of influence of the armed forces. In November 1985, newly elected democratic presidents met to inaugurate a new bridge at Iguaçu Falls on the border between the two nations. They agreed to create a binational commission to study the possibilities of increased economic cooperation and integration. In 1986, the Argentine-Brazilian Program for Integration and Cooperation was announced. In 1986, the two presidents signed the Act of Brazil-Argentine Friendship in Brasília. It was also announced that the two countries would pursue a policy of peaceful nuclear development.

In July 1990, the heads of state of Argentina and Brazil signed the Act of Buenos Aires. The document called for the creation of a common market by 1995, instead of by the earlier announced date of 1999. The Treaty of Asuncíon, signed in March 1991, brought Paraguay and Uruguay into the common market process and marked the creation of a "Common Market of the South," or Mercosur. This new organization's principal goal was to have a free trade area in place among all four countries and a common external tariff (CET) created by 1995. At the time of this writing, the deadline had been extended to 2010. The Protocol of Ouro Preto came into effect in December 1995 and established a new institutional framework for Mercosur. In addition to the Administrative Secretariat located in Montevideo, Uruguay, and the Common Market Council and Common Market Group, set up in 1991, it was decided to add a Joint Parliamentary Commission, a Trade Commission, and a Socio-Economic Advisory Forum. A new date of 2006 was established for the creation of a full common market. As the institutional structure evolved, Chile (1996), Bolivia (1997), Peru (2003), Colombia (2004) and Ecuador (2004) were added as Associate Members (Associate Members enjoy tariff reductions but are not subject to the CET system). In December 2004, Mercosur signed a cooperation agreement with the Andean Community trade bloc, and they published a joint letter of intention for a future negotiation toward integrating all of South America. A year later, President Alvaro Uribe of Colombia signed a law that ratified a Free Trade Agreement with Mercosur. In June 2006, Mercosur expanded further and incorporated Venezuela as a full member.

Tariffs among the four Mercosur member countries were gradually eliminated in the 1990s, and the common market was essentially completed by 1995 with the reduction of 85 percent of tariffs and nontariff trade barriers. Trade expanded rapidly among the Mercosur countries, especially between the two largest countries. According to the IMF's *Trade Statistics Yearbook*, Argentina's exports to Brazil climbed from $3.6 billion to $5.3 billion between 1995 and 1996, and Brazil's exports to Argentina

grew from $4.0 billion to $5.2 billion in the same period.

In retrospect, the late 1990s were the high point of southern cone integration. A maxi-devaluation of the Brazilian currency in January 1999 was the first shock to disturb Mercosur's consolidation. Argentina experienced an institutional crisis in 2001 2 that led to the resignation of President Fernando de la Rua, the collapse of the convertibility plan that had linked the peso to the dollar (one to one), and a massive default on the country's outstanding debt. As a result of the crisis in Argentina, Uruguay's economy stalled and growth rates plummeted and only revived a few years later.

There was little activity in Mercosur in the years following the Argentine-Uruguayan crises. Chile decided that its development goals were best served by negotiating bilateral trade agreements and did so successfully with the United States and China. In an effort to reinvigorate the organization, the heads of state convened the 30th presidential summit, held in Córdoba, Argentina, in July 2006. The formal agenda included discussions of pending trade agreements with external allies, a common customs code, the appointment of members to the Administrative Labor Court, and the establishment of a consistent Mercosur negotiating position for the all-but-collapsed World Trade Organization Doha Round.

Many observers believe that the Córdoba meeting marked the end of Mercosur as it was conceived in 1991. By the end of the two-day meeting, Mercosur appeared to be more of a political organization than a trade facilitating mechanism. The summit addressed the growing importance of political and social integration in addition to strengthening the common market. There were two dramatic aspects to the meeting: the addition of Venezuela as a full member and the attendance of President Fidel Castro of Cuba. The addition of President Hugo Chávez spoke to the increasingly political nature and societal orientation of Mercosur. His public calls to destroy neoliberalism and fight for social justice revealed his ambitions to expand Mercosur's political scope. The

noncommittal responses from the other members, however, suggested internal divisions within Mercosur itself. Nevertheless, for the first time in the history of the common market, the five full members organized a social summit, bringing together civil society leaders who met in a parallel summit to spell out the social needs and goals of the organization.

In a surprise decision, member states addressed the issue of a pending trade agreement between Cuba and Mercosur that would dictate the reduction of tariffs on a number of goods traded between the two entities. Although nominally for economic purposes, the trade agreement emphasized a growing political solidarity that could be interpreted as less than friendly to the United States.

The increasing political nature of Mercosur had been made clear during the Fourth Summit of the Americas in November 2005, held in Mar del Plata, Argentina. At that meeting, the 34 member states of the Organization of American States failed to reach consensus to move ahead with negotiations to conclude the Free Trade Area of the Americas (FTAA). Although 29 states voted to continue the dialogue, with the strong backing of the U.S. delegation led by President George W. Bush, the five dissenting votes cast by the then four full members of Mercosur and by Venezuela were sufficient to destroy the already foundering process.

Although the increasingly political nature of the common market is notable, there are other developments that threaten the future development and consolidation of Mercosur. Paraguay and Uruguay believe that they benefit the least from its current structure. And the reality is that Brazil and Argentina, with more developed and sophisticated industrial agricultural sectors, have benefited disproportionately when compared with the two smaller states.

A 2006 conflict between Argentina and Uruguay highlights the inherent conflict between the two large states and the two smaller partners. Uruguay successfully concluded negotiations to have Spanish and Finnish companies construct two paper mills along the Uruguay River, which forms a natural border between the two states. The mills, estimated at a cost

of $1.6 billion, would represent the largest foreign investment in Uruguay's history. Argentina objected strenuously to the project, arguing that the mills would destroy the river's environment and threaten the fishing and tourism industries. The issue was taken to the International Court of Justice in The Hague; the tribunal's decision favored Uruguay but the government of President Nestor Kirchner in Argentina continued its opposition. As of 2007, it was thought that the principal motivation of the Kirchner government was internal politics there were to be national elections in Argentina in 2007, and this nationalist stance plays well with the voters on the Argentine side of the river.

As of 2007, there were other pending disagreements between the two smaller states and Argentina and Brazil. There was talk in both Paraguay and Uruguay of leaving Mercosur and negotiating trade agreements directly with the United States. Under current rules, Mercosur does not allow bilateral trade agreements outside of the common market. Whether either or both countries would decide to proceed was an important pending issue for the future of the organization.

The institutionalization of Mercosur has been difficult and little had been achieved by late 2007. In this regard, structures were broad but not very deep. The heads of state made all major decisions in periodic summits. The technical staff in Montevideo was small and marginal. The reluctance to give up national sovereignty plagued the organization it was impossible to develop dispute resolution mechanisms, for example, unless member states were willing to cede decision-making authority to supranational tribunals.

There was a growing impression that Mercosur is most valuable to its members as a political counterweight to the United States. That raises doubts about the future. The group dashed the hopes of the United States for a revival of FTAA talks in Mar del Plata. Efforts by the United States to revive the Doha Round discussions met strong resistance from the Mercosur countries, acting within the broader structure of the Group of Twenty (G20), created in 2003 in Cancún, Mexico, as a means of protesting agricultural protection in the European Union and the United States.

Like that of other regional trade movements such as the stalled FTAA, Mercosur's viability has been called into question, and the organization will need to prove it can provide benefits for all member countries to stay intact. As member countries grow disillusioned and increasingly court bilateral trade agreements, Mercosur's unity, influence, and economic potential have diminished. Thus the original goals of Mercosur set forth in the early 1990s have not been achieved, and the organization has taken a polemical turn that deemphasizes integration and gives preference to regional and hemispheric politics.

See also Andean Community; common market; customs unions; Doha Round; Free Trade Area of the Americas (FTAA); regionalism; World Trade Organization

FURTHER READING

Aladi Web site: http://www.aladi.org. Statistics and data bases of trade information for Mercosur member and other Latin American countries.

Baer, Werner, Tiago Cavalcanti, and Peri Silva. 2002. "Economic Integration without Policy Coordination: The Case of Mercosur." *Emerging Markets Review* 3 (3): 269 91. Analyzes the evolution of Mercosur and how the lack of coordination of macroeconomic and ex change rate policies has hampered the realization of re gional trade benefits.

Carranza, Mario E. 2003. "Can Mercosur Survive? Do mestic and International Constraints on Mercosur." *Latin American Politics and Society* 45 (2): 67 103. Ex amines the impact of domestic politics in Argentina and Brazil, and of the relationship between the two countries, on the ability of Mercosur to weather various crises.

Jelin, Elizabeth. 2001. "Cultural Movements and Social Actors in the New Regional Scenarios: The Case of Mercosur." *International Political Science Review* 22 (1): 85 98. Studies how Mercosur's top down regional in tegration policies affect societal and cultural actors, movements, and transformations.

Malamud, Andrés. 2005. "Mercosur Turns 15: Between Rising Rhetoric and Declining Achievement." *Cambridge*

Review of International Affairs 18 (3): 421 36. Provides a retrospective on Mercosur's record of achievements and weaknesses and critically considers prospects for strengthening and enlarging the organization.

Mercosur Web site: http://www.mercosur.int/. Comprehensive archive of official treaties, declarations, and publications of Mercosur, in addition to information about the institutional structure.

RIORDAN ROETT

■ mergers and acquisitions

A merger is the amalgamation of two or more firms into a single firm. This is typically accomplished when one firm acquires the assets of the other firm(s). Such transactions are referred to as mergers and acquisitions (M&As). There are three types: *horizontal* mergers involving firms in the same industry, *vertical* mergers between firms at different stages of the production chain, and *conglomerate* mergers between firms in unrelated industries.

M&As may not involve a complete takeover of another firm's assets. Rather, an investor may acquire a majority or even only a minority stake. This raises the question of what the appropriate ownership structure of the newly created entity should be. Partial ownership entails the sharing of control rights and implies different levels of commitment by the parties involved. There is no clear dividing line between M&As and other forms of organization that allow firms to share assets and coordinate their interests, such as joint ventures.

It is interesting to examine M&As from an international perspective for at least three reasons. First, cross-border M&As have fueled the growth in international production for more than a decade. Since the 1990s, most foreign direct investment (FDI) has been carried out through the acquisition of foreign firms' assets rather than the creation of new firms (the latter also known as greenfield investment). Second, there is evidence that economic integration affects M&A activity by increasing the incentives to undertake cross-border M&As and by forcing industries to

restructure. This restructuring is often accomplished through M&As. Third, both cross-border mergers and mergers between domestic firms engaged in international trade pose challenges for competition policy. Such mergers affect several countries and hence are subject to review by different national competition authorities. These authorities may come to conflicting conclusions, especially if some countries bear more of the costs while others receive more of the benefits of the merger. A prominent example of such conflicts is the merger between two American companies, General Electric and Honeywell, which was approved by U.S. authorities but was ultimately blocked by the European Commission.

Cross-border M&As account for a significant and growing share of total M&A activity. They make up around 25 percent of worldwide M&As. Between 1996 and 2005, the annual average value of cross-border M&As worldwide was U.S. $533 billion, or about 70 percent of annual world FDI flows (UNCTAD 2006). Cross-border M&As, like all M&A activity, tend to occur in waves. In 2000, for example, cross-border M&As peaked at $1.14 trillion. Most cross-border M&As involve developed-country firms. FDI in developing countries is still dominated by greenfield investment. Horizontal mergers account for roughly 70 percent of all cross-border mergers, vertical mergers for around 10 percent. Mergers are much more common in the service sector than in manufacturing (UNCTAD 2000).

Horizontal mergers have attracted the most attention both from scholars and from governmental authorities concerned with competition. The study of these mergers typically involves an analysis of the following two aspects. First, horizontal mergers tend to be anticompetitive. They raise industry concentration and therefore potentially lead to higher prices for consumers or, in the case of intermediate goods, for downstream industries. Another consequence of increased industry concentration is that competitors may benefit even more from the merger than the merging firms themselves. Specifically, as prices rise following the merger, competitors will be tempted to

increase their output, thus gaining market share at the expense of the merging firms.

Second, a merger may allow firms to realize synergies and thus become more efficient. Firms will merge provided that these synergies are sufficiently big to offset the competitive disadvantage outlined in the previous paragraph. If the efficiency gains from the merger are large, prices for consumers may even fall as a result of the merger. But even if they do not, a competition authority may approve a horizontal merger provided that the efficiency gains in production are larger than the losses suffered by consumers due to higher prices. How competition authorities should weigh consumer well-being and economic efficiency in their decisions is an issue that has generated some debate.

The preceding analysis did not distinguish between domestic and cross-border M&As, simply because it applies to both. Still, it is important to point out some of the distinguishing features of cross-border mergers. The main motive for horizontal cross-border M&As is to establish a market presence abroad. They are therefore an alternative to greenfield FDI or exporting. Compared with these alternative choices, cross-border M&As tend to raise industry concentration, thereby potentially hurting consumers and benefiting local competitors. They also offer additional efficiency advantages, however. These include the avoidance of set up costs and of the fixed costs of operating a production facility arising in the case of greenfield FDI, and the avoidance of transportation costs and trade barriers associated with exporting.

Economic integration affects M&A activity in at least two ways. First, it puts pressure on firms to restructure. Specifically, less efficient firms in an industry are forced by import competition to contract or close, whereas improved access to export markets offers more efficient firms an opportunity to expand. An empirically important way in which this restructuring is accomplished is through M&As, whereby efficient firms acquire the assets they need to expand from less efficient firms (Breinlich 2006). Second, trade liberalization raises the incentive to undertake

cross-border mergers. A simple example illustrates the point. Suppose there are two countries, *A* and *B*, each with one firm. When trade barriers are high, there is no incentive for the firms to merge, since each already has a monopoly in its market. As trade barriers fall, competition between the firms increases, giving them an incentive to merge in order to reduce competition.

Except in the European Union, where large cross-border mergers and mergers involving domestic firms with a significant market share in other EU member countries are reviewed by the European Commission, competition policy generally remains the responsibility of national authorities. This may cause problems exactly for the reasons that have led the EU to implement a unionwide competition policy. A simple example helps illustrate them. Consider an industry in which two locally owned firms in country *A* export all their output to country *B*, and suppose there are no local competitors in *B*. A horizontal merger between the two firms allows them to monopolize the market in *B*. This hurts consumers in *B*, but raises profits for the firms in *A*. Synergies involving the elimination of fixed production costs have no effect on prices and hence only boost profits. Provided that competition authorities are guided by considerations of national social welfare, the competition authority in *A* would approve the merger and the one in *B* would prohibit it. This raises the possibility that a merger would be blocked even if the efficiency gains accruing to the firms in *A* outweigh the losses suffered by consumers in *B* so that a merger would raise the overall welfare of the two countries.

The case of the GE/Honeywell merger mentioned earlier is much more complicated than the case illustrated by this simple example. In particular, it has not only a horizontal dimension (both companies produce jet engines for regional aircraft), but also a significant vertical dimension. Specifically, Honeywell is a leading supplier of an intermediate good, namely engine starters, to GE and its two main competitors in the market for large jet engines. A merger between GE and Honeywell may thus have

allowed them to cut off supplies to these two competitors, thereby monopolizing the engine market.

Vertical mergers are best viewed as an alternative to vertical contractual arrangements also known as vertical restraints between firms and their suppliers and/or distributors. An analysis of vertical mergers as in the case of horizontal mergers involves weighing efficiency-enhancing against competition-reducing effects.

See also corporate governance; foreign direct investment (FDI); joint ventures

FURTHER READING

Breinlich, Holger. 2006. "Trade Liberalization and In dustrial Restructuring through Mergers and Acquisi tions." CEP Discussion Paper No. 0717. London: Center for Economic Performance. Available at http:// cep.lse.ac.uk/pubs/download/dp0717.pdf. Shows that the Canada United States Free Trade Agreement led to a significant increase in M&A activity, which transferred substantial resources from less productive to more pro ductive firms.

Horn, Henrik, and Lars Persson. 2001. "The Equilibrium Ownership of an International Oligopoly." *Journal of International Economics* 53 (2): 307 33. Examines how the ownership pattern depends on features of trade and production costs. Shows that private and social incen tives for M&A are aligned for strong synergies, but may differ when the synergies are weak.

Motta, Massimo. 2004. *Competition Policy.* Cambridge: Cambridge University Press. Through case studies compares merger policy in the United States and the European Union.

Neven, Damien J., and Lars Hendrik Röller, eds. 2005. Special Issue on Merger Control in International Mar kets. *International Journal of Industrial Organization* 23 (9 10): 665 848. A collection of articles on merger control.

Raff, Horst, Michael Ryan, and Frank Stähler. 2006. "Firm Productivity and the Foreign Market Entry Decision." Kiel Economics Working Paper 2008 02. Available at http://www.wiso.uni kiel.de/Ordnung+ Wettbewerbspolitik/ewp/EWP 2008 02.pdf. Empirical analysis of different investment options (greenfield FDI, M&A, joint ventures) using Japanese firm level data.

United Nations Conference on Trade and Development (UNCTAD). Various years. *World Investment Report.* New York: United Nations. Annual overview of trends in FDI, including cross border M&As. Detailed report on cross border M&As in 2000.

HORST RAFF

■ migration, international

International migration involves the movement of people, on either a temporary or permanent basis, among the countries of the world economy. Throughout human history these changes of residence have helped to alleviate human suffering, enhance technological progress, and promote cultural exchange. As of 2006 approximately 200 million people, or 3 percent of the world's population, lived outside their country of birth. Although by historical standards this percentage is low, international migration has doubled since 1980. Migration continues to be a key dimension of globalization, albeit one that has complex determinants and outcomes.

As a field of inquiry, international migration has received less attention from social scientists than many other aspects of the world economy. Nonetheless, it is one with a relatively long intellectual history that originally goes back to Ravenstein (1889), with modern contributions going back to Lee (1966). Theoretical and empirical research into international migration underwent a significant renaissance beginning in the 1990s, and multilateral financial institutions gave increased amounts of attention to the subject (e.g., World Bank 2006). A grounded understanding of migration requires an investigation of its recent history, types of migration, the migration decision itself, impacts of migration on both source and destination countries, and current policy debates.

Recent History The modern era of globalization began in the late 19th century. A central component of this was the "Age of Mass Migration" described by Hatton and Williamson (1998). Between 1850 and 1914, approximately 55 million Europeans migrated, most of them unskilled males who settled in

the United States. As Manning (2005) emphasizes, however, the Age of Mass Migration was not just European in nature, with 50 million Chinese and 30 million Indians also migrating (not all voluntarily), primarily to serve as unskilled laborers in British colonies in Africa and the Pacific.

The transatlantic mass migration increased the New World labor force by a third and reduced that of the European economies by an eighth. As a result, the wage gap between North America and the leading European countries narrowed, and flows of migrations from traditional sources slowed. Yet as Western Europe was reaching the end of this migration cycle, other less-developed countries in Eastern Europe were just getting started, giving rise to new sources of migration. The new migration proved to be politically controversial, however, and in 1917 the United States introduced a literacy test for new migrants. This, as well as the outbreak of World War I and its accompanying restrictions on movements of people, effectively brought an end to this era of migration.

Starting in 1917 and continuing throughout the 1920s, isolationist policies characterized much of the Western world. The onset of the Great Depression and the outbreak of World War II only reinforced this trend. International migration declined dramatically. In the aftermath of World War II, however, resurgent nationalism and the spread of communism sent millions of refugees across the European continent. In Western Europe, once a major source of migrants, rapid economic growth in the late 1940s and 1950s led to a shortage of low-wage labor. Initially, the demand was met by migrants from southern European countries, but such sources quickly proved insufficient. By the 1960s, countries across Western Europe were admitting millions of guest workers from Turkey and North Africa. Although the 1973 oil crisis and the ensuing high unemployment brought an abrupt end to these programs, the oil-exporting countries in the Middle East later replicated and expanded the guest-worker model, thereby ensuring that flows of low-skilled migrants remained a significant component of world migration.

In the mid-1960s, Australia, Canada, and the United States overhauled their immigration policies, allowing for a much greater volume of flows and opening the door to migration from countries in Asia, Africa, and Latin America. Together with the precipitous decline in the cost of intercontinental transportation and communication, these reforms have led to steady growth in both the volume and diversity of migrant flows. Yet, as international migration became subject to expanded legal provisions, increasing numbers of migrants opted to forgo legalities altogether. Consequently, undocumented migrants posed increasing challenges for destination countries.

Since the late 1980s, migration flows have continued to grow rapidly. Although Western Europe is an increasingly popular destination for migrants, the United States accepts more immigrants than any other country in the world. The Gulf countries of the Middle East have also emerged as a major destination for migrants, particularly for low-skilled workers from South and Southeast Asia. As of 2006, Mexico and the Philippines were significant sources of low-skilled labor, the latter having 10 percent of its population in foreign countries. China and India predominate as sources of high-skilled migrants.

Types of Migration Migrants leave their home countries and are admitted into destination countries for a wide variety of reasons, and distinct administrative channels have evolved to facilitate these flows. Following are some of the major channels, both legal and unregulated, that define the current international migration system.

Permanent high-skilled migration. Over recent decades, Australia, New Zealand, Canada, and the United States have selectively granted permanent residence to a limited number of high-skilled foreigners who are likely to offer these countries positive economic benefits. Whereas Australia, Canada, and New Zealand operate a "points" system to rate the desirability of potential immigrants, the United States primarily relies on nominations of potential immigrants by local companies that wish to hire them. Among source countries, India and China lead the way, although some high-income countries such

as South Korea and the United Kingdom also supply significant numbers of this kind of migrants. Several Western European countries have recently adopted similar programs to citizenship in order to entice high-skilled migrants.

Temporary high-skilled migration. In many developed countries, programs that grant permanent residence to foreigners who do not have ties to the destination country can be politically controversial. In such cases, governments may seek to fill occupational shortages through the recruitment of high-skilled migrants on a temporary basis. Historically, these flows have been concentrated in education and health-related services. During the 1990s, however, booms in information and communications technology (ICT) led to a shortage of related skills in many high-income countries, resulting in a jump in flows of technology professionals, mostly from India. In the early years of the 21st century, even nontraditional immigration destination countries, such as China and the Czech Republic, began to grant temporary work visas to high-skilled foreigners.

Temporary low-skilled migration. Despite the fast growth of temporary high-skilled migration, these flows are dwarfed by temporary low-skilled migration, in which countries admit migrant workers to provide low-cost services on a strictly temporary basis. Countries typically implement these programs when rapid economic growth has improved the wages and work conditions of the local workforce and left them correspondingly unwilling to work at low-wage jobs. One of the best-known programs of this kind was the West German *Gastarbeiter* program of the 1960s and 1970s in which West Germany recruited large numbers of Turkish and North African guest workers. These programs were rolled back in the low-growth, high-inflation period of the 1970s, but low-skilled migration programs have since grown significantly in the Middle East. In 2006, approximately 10 million temporary low-skilled migrants were employed in the Saudi Arabia, Kuwait, and the UAE. India and Pakistan are major sources of manual laborers and construction workers, with domestics, nurses, and other service workers coming primarily from Sri Lanka, the Philippines, and Thailand. The

rights of temporary low-skilled migrants, however, are often not well defined or protected.

Family migration. Family migration is among the largest official channels of migration and represents a disproportionate share of flows from low- and middle-income countries to high-income countries. This mode of migration enables foreign spouses of citizens, children born abroad, and even foreign-born parents and siblings of citizens to gain permanent residency in the country of the respective relative. Although almost all countries in the world permit some form of family migration, countries differ significantly in their definitions of which types of relatives are eligible to migrate, and which are not. Over recent years, many high-income countries have tightened restrictions on family migration in order to limit the volume of politically controversial inflows of migrants from low-income countries.

Coethnic and national priority migration. A number of programs provide permanent residency to foreigners on the basis of their ethnic background, religious affiliation, or national origin. In many European countries, coethnic migration functions through the liberal application of the *jus sanguinis* citizenship tradition, whereby citizenship rights can be inherited through long lines of ancestry. Israel's Law of Return is one of the most controversial instruments of coethnic and national priority migration as it provides a path to permanent residency to all those who practice Judaism, irrespective of any other criteria. The United States operates dedicated paths to permanent residency for migrants from Cuba and a number of other countries, while also operating a "diversity visa" lottery of the United States, purportedly to diversify immigration flows, but which in fact was originally proposed by lawmakers as a means to provide a new channel for Irish migration to the United States.

Asylum seekers. The 1951 Geneva Convention protects persons with a "well founded fear of persecution [by state agents] for reasons of race, religion, nationality, membership of a particular social group, or political opinion" from return to their country of origin. Asylum seekers commonly invoke the 1951 convention in seeking permission to remain perma-

nently in a destination country. In order to assess asylum claims, major recipient countries have established elaborate administrative legal procedures, including access to counsel and the right to appeal initial decisions. As a result, individual claims can sometimes take years to resolve. In some destination countries, asylum seekers qualify for social services. In others, they are granted only the right to work. In yet others, they remain in detention until a final decision on their application is reached.

Refugees. Groups of people who have fled to neighboring countries due to war, famine, environmental collapse, or political strife are considered refugees. Refugees are often hosted in makeshift camps set up by international humanitarian agencies along the border of the affected country with the expectation that they will return to their country of origin at the conclusion of the disturbance. Where this is not possible, the United Nations High Commissioner for Refugees (UNHCR) might seek to resettle refugees in a third country or help them to settle in the country of refuge.

Undocumented migration. Reliable data on the number of people involved in undocumented migration are limited and estimates vary widely. There is evidence, however, that flows increased markedly in the 1990s and the early years of the 21st century, particularly to Europe, where the implementation of the free movement of labor within the European Union has increased the potential rewards for illegal migrants who reach a member country. Major flows also exist between Mexico and the United States and among countries in West Africa, Southern Africa, and Southeast Asia. Flows of undocumented migrants include those who move voluntarily and those who are trafficked against their will.

Visa-free migration. Visa-free migration exists (with some exceptions) within the European Union, as well as between New Zealand and Australia. This channel grants citizens the right to work for an unlimited time in any of the countries that are party to the agreement.

The Migration Decision The economic, social, and psychological consequences of migration to the individual are profound, and it can be very difficult

for anyone to predict with any certainty whether migration will actually improve their life. The inherent unpredictability of migration tends to discourage people from leaving their home countries, even when doing so would increase their expected income. Those who do choose to migrate are often unusual in their willingness to tolerate risk. The predominance of young males in migration flows reflects this reality.

Migration is an expensive endeavor in terms of both the direct costs involved and the opportunity costs of leaving a livelihood in the home country. These costs prevent a large number of people from migrating even if they wish to do so. This, in conjunction with the unwillingness of those earning high incomes to migrate, gives rise to the so-called migration hump. That is, those with the highest inclination to migrate come from middle-income countries, where wages are high enough to provide the base level of wealth necessary to finance migration, but also low enough to generate significant financial incentives for migrating to high-income countries. Thus as incomes in middle-income countries increase, migration rates tend to decline. On the other hand, when incomes rise in low-income countries, migration rates tend to increase. Such was the case during the Age of Mass Migration, with significant increases in emigration rates following industrial revolutions by four or more decades.

Another important factor that can either spur or restrict migration is the presence of familial or trusted contacts in the destination country. If a large number of people from their community have migrated, potential migrants have easy access to information on the experiences of other migrants and relatively high confidence that their migration experience will conform to what they are told. Immigrant communities can also provide essential services to new arrivals, such as accommodation and employment, thereby lowering the cost and risk of resettlement.

In some cases, neither financial resources nor information networks in destination countries allow migrants to fully assess their prospects. Some high-skilled migrants, for example, fail to secure a job that

utilizes their skills. This causes what analysts refer to as brain waste.

Impacts on Source Countries High-skilled migrants are commonly trained at substantial costs to the taxpayers of source countries through public education systems. Their departure thus has profound effects in the form of what is known as *brain drain*. Source countries can also lose tax revenues that migrants would have generated. More important, many of the skills sent from less-developed to more-developed countries are already scarce in source countries. In the case of medical services, for which more-developed countries have a strong desire and less-developed countries an urgent need, brain drain can potentially imperil the quality of medical care. In many southern African countries, high emigration rates for health care professionals have existed alongside unprecedented medical crises brought on by the HIV/AIDS pandemic. In Malawi, for example, it is estimated that approximately half of the country's nursing staff has been lost to migration. Meanwhile, the rate at which Malawian women die during pregnancy and childbirth doubled during the 1990s to one of the highest levels in the world.

The emigration of skilled workers does not always create problems for source countries. In some cases, emigration alerts outside investors to a large or relatively underused skill base of the source country. The success of skilled Indian migrants in the United States, for instance, possibly helped to spur the large inflow of ICT-related foreign direct investment (FDI) to India seen over recent years. Thus, when the conditions are right, skilled migrants are able to generate networks of investment, trade, and technology transfer that increase the productivity and demand for skills in the home country while extending the global technology frontier and lowering the cost of products used by billions of people worldwide.

Another potentially compensating benefit of the brain drain is that it tends to increase the demand for skills in the source country by raising the rate of return to education. Some researchers have suggested that, even accounting for the emigration of skilled individuals, the increase in demand for education generated by the brain drain may actually *increase* the number of skilled workers in the population. This is known as a brain gain. Although brain gain outcomes are possible, they depend on large responses in the supply of education and training. They are not, therefore, a *general* outcome of high-skilled migration.

The most easily quantifiable benefit of emigration to source countries is the flow of money, or *remittances*, sent by migrant workers to their home countries. Recent estimates suggest that the total remittance flow to developing countries exceeds U.S. $200 billion. In a number of countries, remittance inflows are larger than inflows of FDI and can compose up to 10 percent of national incomes. Such flows can make a significant difference for families living in poverty in source countries, which is a common reason for communities to allow and sometimes even encourage their family members to seek work abroad.

Under the auspices of the World Trade Organization (WTO), the liberalization of services trade has occurred in a number of sectors, such as finance and telecommunications, of interest to developed countries. The WTO's General Agreement on Trade in Services recognizes the *temporary movement of natural persons* as a way to export certain labor-intensive services such as housekeeping and construction. Given the natural comparative advantage of developing countries in such labor-intensive services, this channel could be of great importance to their trade and development prospects. The WTO protocol on the temporary movement of natural persons, however, is largely limited to the exchange of corporate personnel and is not designed to enhance the delivery of labor-intensive services.

Impacts on Destination Countries The world's most productive economies often depend on and benefit greatly from the presence of migrant workers. Currently, the world's richest countries continue to import labor at high rates. For example, in Southeast Asia's richest country, Singapore, migrants make up around one-quarter of the workforce. In Europe, those countries with the highest number of migrant workers, such as Switzerland and Luxembourg, are

also the wealthiest. In Dubai, currently among the world's fastest expanding areas of economic activity, as of 2006 there were nine times as many migrant workers as resident nationals.

Across all regions of the world, the correlation between levels of economic growth and immigration is clear, not only because migrants wish to move to where they can earn the highest wages, but also because immigration is often much less politically controversial when an economy is growing quickly and unemployment is low, thereby easing the liberalization of migration restrictions.

Critics of immigration are usually concerned with its impact on social fabric, national culture, environment, social welfare programs, or the wages of low-skilled workers. Correspondingly, advocates of increased immigration have turned their attention to the potential role further inflows could play in ensuring the sustainability of pension systems in the face of aging populations.

A common claim against migration is that migrants, particularly those who are low-skilled or undocumented, consume far more in public services than they contribute in tax revenue to destination countries. However, many studies paint a different picture. Research in Europe, for instance, indicates that immigrants (even undocumented immigrants) contribute more in taxes and pension funds than they consume in benefits or other public services. Interestingly, some commentators have argued that increased migration will actually be critical to the survival of social security systems in high-income countries over the coming decades. For instance, as the proportion of retirees in the population increases, many social security systems will need to find sources of tax revenues in order to fund pension payouts. Migrants are one potential source through which tax revenues might be increased.

The contention that migrants "take the jobs" or reduce the wages of native workers is probably the most common argument advanced against migration. Economic theory suggests that increasing the size of the labor force will lower the wage level, all other things being equal. Yet when a region receives an influx of immigrant labor, native workers may

choose to move to other regions, thereby masking any changes in the aggregate wage. Firms may also decide to move into a region experiencing such an influx, causing local wages to increase. Accordingly, it has been very difficult for studies to accurately assess the impact of migration on unemployment or wage levels and fierce academic debates continue to rage on the topic.

Unlike statistically testable hypotheses regarding the effect of migration on public finances or low-skill wages, claims that immigrants alter culture or threaten national cohesion are difficult to assess. Fears that migrants threaten the cultural or social order remain among the most instrumental sources of resistance to migrant inflows. Yet the historical record shows that many of the perceived cultural idiosyncrasies of migrants fail to outlive the first generation, while others, such as music, cuisine, and sport, are co-opted by the native culture due to their popularity among natives.

Policy Priorities Migration has always been and will continue to be a central feature of the world economy. It does, however, involve a number of important policy issues that have not yet been resolved: increased multilateral coordination of migration flows, prevention of human rights abuses, care of refugee populations, facilitation of the temporary movement of natural persons in services trade, the harnessing of remittance flows for poverty-reducing investments, management of brain drain, and reduction of brain waste. These priorities compose a large and important policy agenda for better harnessing this important socioeconomic process for development purposes in the modern world economy.

See also brain drain; brain gain; brain waste; European Union; General Agreement on Trade in Services (GATS); information and communications technology; migration governance; remittances; temporary movement of natural persons

FURTHER READING

Adams, Richard H. Jr., and John Page. 2005. "Do International Migration and Remittances Reduce Poverty in Developing Countries?" *World Development*

33 (10): 1645 69. An empirical examination of the in
fluence of migration and remittances on poverty levels
in a relatively large sample of developing countries.

Borjas, George J. 1994. "The Economics of Immigration."
Journal of Economic Literature 32 (4): 1667 1717. An
often cited, early review of the research on international
migration.

Chiquiar, Daniel, and Gordon H. Hanson. 2005. "Inter
national Migration, Self Selection, and the Distribution
of Wages: Evidence from Mexico and the United
States." *Journal of Political Economy* 113 (2): 239 81. An
empirical examination of the role of skill in international
migration decisions.

Global Commission on International Migration. 2005.
*Migration in an Interconnected World: New Directions
for Action.* Geneva. A thorough overview of interna
tional migration trends and policies that summarizes the
work of a commission of notable international policy
makers.

Hatton, Timothy J., and Jeffrey G. Williamson. 1998. *The
Age of Mass Migration: Causes and Economic Impact.*
Oxford: Oxford University Press. A central, historical
reference on the history of international migration.

Lee, Everett S. 1966. "A Theory of Migration." *Demography*
3 (1): 47 57. An early theoretical analysis that elucidates
some basic features of migration.

Manning, Patrick. 2005. *Migration in World History.* Lon
don: Routledge. A comprehensive, historical account of
migration going back to the movements of the earliest
humans.

Mora, Jorge, and J. Edward Taylor. 2006. "Determinants of
Migration, Destination, and Sector Choice: Disen
tangling Individual, Household, and Community Ef
fects." In *International Migration, Remittances and the
Brain Drain*, edited by Çaglar Özden and Maurice
Schiff. Washington, DC: World Bank and Palgrave
Macmillan, 21 51. An important empirical exploration
of the variables that help to explain the migration deci
sion.

Nyberg Sorensen, Ninna, Nicholas Van Hear, and Poul
Enberg Pedersen. 2002. *The Migration Development
Nexus: Evidence and Policy Options.* Geneva: Interna
tional Organization for Migration. A thorough review of
the role of migration in development with a helpful set of
references for researchers.

Ottaviano, Gianmarco I. P., and Giovanni Peri. 2005.
"Rethinking the Gains from Immigration: Theory and
Evidence from the U.S." NBER Working Paper 11162.
Cambridge, MA: National Bureau of Economic Re
search. A sophisticated empirical examination of the way
that immigrant and native born workers substitute for
one another in U.S. labor markets.

Ravenstein, Ernest George. 1889. "The Laws of Migration."
Journal of the Royal Statistical Society 52 (2): 241 305.
One of the earliest systematic inquiries into the phe
nomenon of migration.

World Bank. 2006. *Global Economic Prospects 2006: Eco
nomic Implications of Remittances and Migration.* Wash
ington, DC. An important look at migration from a
development perspective that provides a number of
policy recommendations.

ANDREW L. BEATH, IAN GOLDIN,
AND KENNETH A. REINERT

■ migration governance

People migrate from one nation state to another, and
governance of international migration refers to the
national and international laws and norms regulating
such movements. Even though international migra
tion by definition involves two nation states, and the
people moving between them, most migration oc
curs under the national laws of receiving states, al
though these national laws may be shaped by inter
national norms.

The best example of international governance of
migration applies to refugees, about 5 percent of the
world's 200 million migrants. The 1951 Geneva
Convention relating to the Status of Refugees and its
1967 Protocol define a refugee as a person who,
"owing to well-founded fear of being persecuted for
reasons of race, religion, nationality, membership of
a particular social group or political opinion, is out
side the country of his nationality and is unable or,
owing to such fear, is unwilling to avail himself of the
protection of that country." The 138 signatories to
the Convention and Protocol pledge not to return or
"refoul" persons who satisfy this definition to their
countries of origin.

Geneva Convention countries sometimes resettle persons recognized by the UN High Commissioner for Refugees (UNHCR), as when the United States resettled Vietnamese who fled to Thailand and other Southeast Asian countries in the late 1970s. Other countries that do not resettle refugees often participate in burden sharing, providing funds to support refugees who have taken refuge in poor countries, such as Afghanis in Pakistan, and supporting repatriation when their countries of origin become safe.

Countries also establish systems to screen foreigners who arrive and apply for asylum, with those who need protection being allowed to stay. Countries differ in how they implement the Geneva refugee convention, with traditional immigration receiving countries such as the United States and Canada resettling more refugees, while European countries such as Britain and France deal with asylum seekers. National laws determine exactly what assistance refugees and asylum applicants receive. European countries often provide housing and other assistance to both refugees and applicants for asylum, while the United States relies on private organizations such as churches to integrate refugees and care for asylum applicants.

Governance and Migration The 1995 report of the Commission on Global Governance defined governance as "the sum of the many ways individuals and institutions, public and private, manage their common affairs . . . a continuing process through which conflicting or diverse interests may be accommodated and cooperative action taken." Governance of international issues is agreed to by nation-states that see advantages in creating rules and norms, and institutions to ensure that they are followed. For example, the World Trade Organization (WTO) establishes rules for international trade in goods and services, establishing mechanisms to resolve trade disputes, and the International Labor Organization (ILO) promotes policies that protect workers. Nation-states may delegate a part of their national sovereignty to international institutions such as the WTO and ILO, agreeing to incorporate the agreements or conventions they establish into national laws governing trade and worker rights.

International organizations such as WTO and ILO set rules for the behavior of nation-states, while organizations such as the International Organization for Migration (IOM) provide services to states. IOM began as an intergovernmental organization that moved refugees and displaced persons to new homes at the end of World War II. It has evolved into an organization of 118 countries that aims to improve migration management by providing services as well as advice to governments. IOM acts as secretary for a variety of ad hoc efforts to improve migration management, including the Berne Initiative that led to the International Agenda for Migration Management, a "non-binding reference system and policy framework."

The stock of international migrants doubled between 1985 and 2005 to almost 200 million; migrants are persons outside their country of citizenship a year or more, regardless of their legal status and reason for being abroad. International migration is likely to continue increasing because of demographic and economic inequalities among countries as well as revolutions in communication and transportation that make it easier to learn about opportunities abroad and travel over national borders to take advantage of them. Winning international agreement on a system to manage migration is difficult, however, because sovereignty includes the right to determine who enters and stays in a country. With no consensus on whether migration is good or bad for sending and receiving countries, there is no legal and institutional framework for dealing with migration cooperatively on a global scale except for UNHCR and the protection of refugees.

Developing a global policy framework is difficult because policy contradictions at the national level are mirrored at the international level. This occurs in other areas with global frameworks, as when national governments promote freer trade at the WTO but protect their farmers with subsidies and import barriers, making it hard for a WTO that operates by achieving consensus among its nation-state members to lower farm trade barriers. Similarly, if industrial countries want to encourage the best and brightest from developing countries to work and settle, they

may retard economic development and accelerate unwanted migration of less-skilled workers, so that global discussions of such migration can quickly devolve into arguments over compensation for the brain drain.

Top Down, Bottom Up Most international organizations operate in a top-down fashion, seeking to establish norms that protect migrants in nation-states that ratify conventions and protocols approved by the organization's members. International human rights conventions give individuals the right to leave their country, but there is no comparable right to enter or stay in another country. The ILO approved major conventions in 1949 and 1975 to protect migrant workers, and the United Nations General Assembly approved a more expansive migrant convention in 1990. Finally, the UN approved two protocols in 1998 aimed at curbing smuggling and trafficking in persons over national borders.

The ILO's Convention 97 (1949) established the fundamental principle of equality of treatment for migrant workers, meaning that migrants should be treated as other workers in the countries in which they work. The ILO called for migrant workers to move across national borders under the terms of bilateral agreements and to have labor rights ranging from the right to organize unions to the right to work-related benefits. ILO Convention 143 (1975) emphasized the steps governments should take to minimize illegal migration and to promote the integration of settled migrants.

Both of these controversial provisions, ILO migrant conventions have fewer-than-average ratifications, 42 for Convention 97 and 18 for Convention 143 as of 2005. The reluctance of countries to ratify ILO migrant conventions is often attributed to provisions in the conventions that conflict with national legislation. For example, migrant workers in the United States have the same union rights as U.S. workers, but if unauthorized migrants are unlawfully fired because of their union activities, they are not entitled to pay for the time they did not work. The U.S. Supreme Court held in Hoffman Plastics (2002) that requiring back pay for unauthorized migrants would "encourage the successful evasion of

apprehension by immigration authorities, condone prior violations of the immigration laws, and encourage future violations." In effect, the court ruled that an unauthorized worker's violation of immigration laws was more serious than an employer's violation of labor laws.

On December 18, 1990, the UN General Assembly approved the International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families. This 8-part, 93-article convention went into force in July 2003 to "contribute to the harmonization of the attitudes of States through the acceptance of basic principles concerning the treatment of migrant workers and members of their families." The UN Convention has been ratified by 35 major emigration sending countries, in part because it goes beyond the protections of ILO Conventions to cover all migrants, authorized and unauthorized.

The major employment-related protections of the UN Convention are in part III, particularly articles 25 27, which prescribe equality in wages and working conditions for authorized and unauthorized migrant and national workers, assert that migrants should be allowed to join unions, and call for migrant workers to receive benefits under social security systems to which they contribute, or to receive refunds of their social security contributions on departure. Authorized migrants should have additional rights set out in part IV, including the right to information about jobs abroad before they arrive as well as a list of "equal treatments," including freedom of movement within the host country, freedom to form unions and participate in the political life of the host country, and equal access to employment services, public housing, and educational institutions.

The United Nations has a High Commissioner for Human Rights that examines compliance with seven "core human rights treaties" that implement the Universal Declaration of Human Rights (1948). These include the International Convention on the Elimination of All Forms of Racial Discrimination (1965); the International Covenant on Civil and Political Rights (1966); the International Covenant on Economic, Social, and Cultural Rights (1966);

the Convention on the Elimination of all Forms of Discrimination against Women (1979); the Convention against Torture and Other Cruel, Inhuman, or Degrading Treatment or Punishment (1984); the Convention on the Rights of the Child (1989); and the International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families (1990).

Other international instruments and declarations also call for equal treatment for migrants. The Vienna Declaration and Programme of Action on Human Rights (1993) and the Cairo Programme of Action of the International Conference on Population and Development (1994) affirmed the importance of promoting and protecting the human rights of migrant workers and their families, and the Beijing Platform of Action of the Fourth World Conference on Women (1995) paid special attention to the rights of women migrants and urged that migrants be protected from violence and exploitation. The World Conference on Racism, Racial Discrimination, Xenophobia, and Related Intolerance in 2001 issued the Durban Declaration and Program of Action, calling on countries to allow migrants to unite their families and to make active efforts to reduce discrimination against migrant workers. The UN General Assembly in 2000 adopted the Convention against Transnational Organized Crime, which has two additional protocols: the Protocol to Prevent, Suppress, and Punish Trafficking in Persons, especially Women and Children, and the Protocol against the Smuggling of Migrants by Land, Sea, and Air.

The UN in 2006 held a high-level dialogue on migration and development, endorsing a new "moving forum" to discuss best practices to maximize the development impacts of migration, remittances, and migrants' returning or staying abroad but forging new trade and investment links to their countries of origin. The forum is an alternative to proposals for a new "World Migration Organization" analogue to the WTO, advocated by some as a way to improve the governance of international migration.

Most migration is governed unilaterally by receiving countries, which decide whether they are destinations for immigrants, guest workers, and other types of foreigners and establish rules governing entries and stays. Since most migration occurs between neighboring countries, however, there are many bilateral and regional agreements dealing with migration. The world's largest bilateral migration flow involves an average 200,000 Mexican immigrants, millions of nonimmigrant visitors and guest workers, and the settlement of perhaps 400,000 unauthorized Mexicans each year in the United States. There are regular consultations between Mexican and U.S. immigration officials, and former Mexican president Vicente Fox (2000 2006) made improving the status of Mexicans in the United States his top foreign policy priority.

The Regional Migration Conference, or the Puebla Process, an initiative launched by the Mexican government in the city of Puebla in 1996 in response to voter approval of Proposition 187 in California in 1994, includes 11 countries that meet at least once a year to discuss migration issues: Canada, the United States, Mexico, and Central American countries plus the Dominican Republic. The discussions cover changes in national migration policies, the link between migration and development, migrant trafficking, cooperation for the return of extraregional migrants, and the human rights of migrants.

Puebla Process consultations are credited with paving the way for the United States to legalize the status of many Central Americans who fled to the United States during civil wars in the 1980s, and to grant temporary protected status to Central Americans in the United States when Hurricane Mitch carved a path of destruction in 1998 and after El Salvador had severe earthquakes in 2001, and with encouraging cooperation to improve safety at the Mexican-U.S. and Mexican-Guatemalan borders. There are many other regional migration forums, including those between the European Union and North African countries, among African countries, and among the Andean countries of South America. These forums discuss economic issues such as remittances and development as well as migration management issues such as traffickers and criminals (International Organization for Migration 2005b).

Whither Migration Governance About two-thirds of the world's migrants are in the high-income countries that have one-sixth of the world's people but five-sixths of the world's economic output. Even though there are almost as many migrant workers from developing countries in other developing countries (about 30 million in 2005) as there are developing country migrants in industrial countries (31 million), most unresolved governance issues involve developing and industrial countries. For example, one major issue is conditionality, or whether industrial countries that provide aid or open doors for guest workers can require sending countries to cooperate to accept the return of apprehended nationals.

Many observers believe that bringing governments together to talk about migration is the first step toward improved governance. Major concrete outcomes of regional organizations aimed at improving migration governance include making it more difficult for unauthorized foreigners to apply for asylum in more than one country. Both the United States and Europe have persuaded neighboring southern countries to prevent the transit of migrants headed for their borders (pushing the borders out), sometimes with financial assistance, as when the United States helps Mexico to prevent the transit of third-country nationals headed for the United States.

Most migration occurs within regions, and regional organizations that discuss migration issues can be more informal and flexible than global organizations that fix rules and enforcement mechanisms. Most regional organizations include sending and receiving countries but do not develop binding rules. It is generally far easier to achieve consensus in a closed-door meeting on regional migration issues than to negotiate global rules in forums in which national interests may vary widely.

The role of migration governance in the modern world economy is to establish rules governing the movement of workers over borders for temporary periods of employment. The fundamental principle in international conventions protecting migrants is equality, treating migrant workers as local workers. The gap between this equality goal and migrant realities is often large.

See also brain drain; brain waste; International Labor Organization; migration, international; remittances; temporary movement of natural persons

FURTHER READING

Gibney, Matthew, and Randall Hansen, eds. 2005. *Immigration and Asylum: From 1900 to the Present*. Santa Barbara, CA: ABC CLIO. Downloadable from http://www.abc clio.com/products/overview.aspx?productid=108786. Covers the major types of migration and migrants, with extensive references to international, regional, and national migratory movements.

International Organization for Migration. 2005a. "Costs and Benefits of International Migration." Download able from http://www.iom.ch/jahia/Jahia/cache/bypass/pid/8?entryId=932. A reference that provides perspectives on a major migration issue on a biennial basis.

———. 2005b. "Interstate Cooperation and Migration." Downloadable from http://www.iom.ch/jahia/Jahia/cache/bypass/pid/8?entryId=932. Includes contributions from governments and academics discussing the challenges and opportunities of cooperating to improve migration management.

Martin, Philip, Manolo Abella, and Christiane Kuptsch. 2006. *Managing Labor Migration in the Twenty First Century*. New Haven, CT: Yale University Press. Overview of labor migration issues that includes a global survey of stocks and flows and issues in various regions and countries, closer examination of migration at the top and bottom of the job ladder, and a comparison of migration and other flows, including foreign direct investment and aid.

Martin, Philip, Susan Martin, and Patrick Weil. 2006. *Managing Migration: The Promise of Cooperation*. Lanham, MD: Lexington Books. Provides an overview of migration management issues and discussions of specific migration corridors, including Mexico United States, Morocco Spain, and Albania Italy.

Migration News. Quarterly. Downloadable from http://migration.ucdavis.edu/ Comprehensive analysis of on

going migration developments in the Americas, Europe, Asia, and other regions of the world.

PHILIP MARTIN

■ Millennium Development Goals

The Millennium Development Goals (MDGs) were introduced in the Millennium Declaration, which was signed by 189 states and adopted by the General Assembly of the United Nations during the United Nation's Millennium Summit on September 6 8, 2000. The 8 MDGs and 18 targets formulate a comprehensive set of development objectives to be attained by 2015. Additionally, the United Nations observes 48 indicators to measure progress toward the MDGs. As the MDGs have been endorsed by a large number of UN member governments and international organizations, they also provide a framework of accountability for the outcomes of international development efforts.

Specifically, the Millennium Declaration establishes the following goals and targets:

Goal 1: Eradicate extreme poverty and hunger.
  Target 1: Halve, between 1990 and 2015, the proportion of people whose income is less than $1 a day.
  Target 2: Halve, between 1990 and 2015, the proportion of people who suffer from hunger.
Goal 2: Achieve universal primary education.
  Target 3: Ensure that by 2015, children everywhere, boys and girls alike, will be able to complete a full course of primary schooling.
Goal 3: Promote gender equality and empower women.
  Target 4: Eliminate gender disparity in primary and secondary education, preferably by 2005, and at all levels of education no later than 2015.
Goal 4: Reduce child mortality.
  Target 5: Reduce by two-thirds, between 1990 and 2015, the mortality rate among children younger than five.

Goal 5: Improve maternal health.
  Target 6: Reduce by three-quarters, between 1990 and 2015, the maternal mortality rate.
Goal 6: Combat HIV/AIDS, malaria, and other diseases.
  Target 7: Have halted by 2015 and begun to reverse the spread of HIV/AIDS.
  Target 8: Have halted by 2015 and begun to reverse the incidence of malaria and other major diseases.
Goal 7: Ensure environmental sustainability.
  Target 9: Integrate the principles of sustainable development into country policies and programs and reverse the loss of environmental resources.
  Target 10: Halve by 2015 the proportion of people without sustainable access to safe drinking water and basic sanitation.
  Target 11: Have achieved a significant improvement by 2020 in the lives of at least 100 million slum dwellers.
Goal 8: Develop a global partnership for development.
  Target 12: Develop further an open, rule-based, predictable, nondiscriminatory trading and financial system (including a commitment to good governance, development, and poverty reduction, nationally and internationally).
  Target 13: Address the special needs of the least-developed countries (including tariff- and quota-free access for exports of the least-developed countries; enhanced debt relief for heavily indebted poor countries and cancellation of official bilateral debt; and more generous official development assistance for countries committed to reducing poverty).
  Target 14: Address the special needs of landlocked countries and small island developing states.
  Target 15: Deal comprehensively with the debt problems of developing countries through national and international

measures to make debt sustainable in the long term.

Target 16: In cooperation with developing countries, develop and implement strategies for decent and productive work for youth.

Target 17: In cooperation with pharmaceutical companies, provide affordable access to essential drugs in developing countries.

Target 18: In cooperation with the private sector, make available the benefits of new technologies, especially information and communication.

The United Nations measures progress toward these goals and targets using 48 Millennium Development Indicators, such as the proportion of people living on less than U.S. $1 per day, the net enrollment rate in primary education, the ratio of literate women to men (ages 15 to 24), the mortality rate of children younger than five, HIV prevalence among women ages 15 to 24, the proportion of land area covered by forest, and the level of official development assistance.

As of mid-2007, progress toward achieving the MDGs was uneven. For example, the World Bank's *Global Monitoring Report 2006* observed that the share of people in developing countries living on less than U.S.$1 per day was expected to fall to 10.2 percent by 2015, from 27.9 percent in 1990. Most of this progress was accounted for by large declines in poverty rates in Asia, while poverty rates in sub-Saharan Africa had improved only modestly since 1990, from 44.6 percent in 1990 to 44.0 percent in 2002, and under 2007 projections were expected to fall to 38 percent by 2015, far above the target of 22.3 percent.

By specifying quantitative targets for development outcomes, the MDGs also provide a framework for holding development agencies, international organizations, and recipient countries accountable for the outcomes of development assistance. For example, they lend themselves to analyses of whether aid commitments are consistent with or fall short of the goal of attaining the MDGs. (The *Global Monitoring Report 2006* suggests that, in most cases, aid commitments fall short of the goal.)

The commitment to attaining the MDGs has contributed to (or at least coincided with) an increase in official development assistance, which has risen from an average of U.S. $59 billion annually (1996 2000) to U.S. $106 billion in 2005 for Organisation for Economic Co-operation and Development (OECD) donor/creditor countries, partly reflecting an increase in debt relief and the emerging international response to HIV/AIDS. Although the 2006 data on external aid showed no further increase in external aid (excluding debt relief) for OECD countries, aid specifically to sub-Saharan Africa was expected to increase, in line with commitments made at the G8 Summit in Gleneagles in 2005, and some "new" donor countries, most notably China, had increased their aid commitments.

Have the MDGs been effective? Although it is too early (as this is written in 2007) to answer this question, recent experience provides some pointers. Overall aid has increased during the period the MDGs relate to, although these increased efforts fall short of what would be required to meet the MDGs. Nevertheless, the MDG targets provide a measure of accountability for aid outcomes, and therefore reinforce or sustain the momentum toward increased development aid.

See also aid, international; United Nations Conference on Trade and Development

FURTHER READING

Most of the information relating to the Millennium Development Goals is available online, including through Web sites that provide regular data updates. The most important ones are the United Nations Millennium Assembly Web site (http://www.un.org/millennium/), the United Nations Millennium Goals Web site (http://www.un.org/millenniumgoals/), the United Nations Millennium Development Indicators Web site (http://mdgs.un.org/), and the World Bank's MDG Web site (http://ddp ext.worldbank.org/ext/GMIS/home.do?siteId=2).

MARKUS HAACKER

■ monetary conditions index

A monetary conditions index (MCI) is a simple device that combines movements in different financial variables, notably the interest rate and the exchange rate, into a single number. It serves as an easy-to-understand reference or information variable to financial markets and the general public at a point in time relative to some point in the past. In addition, a MCI can function as a short-term operating target in the conduct of monetary policy in small open economies.

The Bank of Canada pioneered the construction of a MCI in the early 1990s. The original MCI consisted of a weighted average of the real short-term interest rate and the real exchange rate, the respective nominal rate adjusted for inflation and the price of domestic relative to foreign goods. In practice, however, the MCI has been stated in terms of the nominal short-term interest rate and the nominal exchange rate, two of the most readily observable financial variables in the economy.

The basic idea behind the MCI is that in an open economy, both the real interest rate and the real exchange rate help determine aggregate demand. They can be combined to produce a simple gauge that measures the influence of both variables on aggregate demand. This indicator captures changes in the stance of monetary policy and the associated effects on the exchange rate. In addition, it records changes in the exchange rate that are not accompanied by changes in the interest rate, such as terms of trade shocks or loss of confidence. Finally, it registers movements of the interest rate and exchange rate in opposite directions in situations where, for instance, an announced policy change lacks credibility.

Explaining the MCI In its simplest form an aggregate demand relation can be modeled as follows:

$$y_t = -a_1 r_t - a_2 e_t + v_t \qquad a_1, a_2 > 0$$

where

y_t = aggregate demand,

r_t = real interest rate,

e_t = real exchange rate (rise implies that domestic currency is appreciating, all other things being equal), and

v_t = other factors that influence aggregate demand.

The size of a_1 and a_2 reflect the relative effect of the real interest rate and exchange rate channel on aggregate demand. Both parameters are important ingredients in the construction of the MCI:

$$MCI_t = (r_t - r_0) + \frac{a_2}{a_1}(e_t - e_0) + 100$$

The MCI at time t is a weighted sum of the change in the real rate of interest and the change in the real exchange rate relative to the base period. The interest rate is measured in percentage points while the exchange rate appears in index form with 100 being its value in the base period. The selection of the base period is arbitrary. The weight on the real exchange rate is typically less than or equal to one.

According to the above specification, a one-point rise in the real interest rate at time t is associated with a one-point rise in the MCI. If the relative weight on the real exchange rate is one half, then a two-point rise in the real exchange rate also leads to a one-point increase in the MCI. It is important to keep in mind that the absolute level of the MCI is meaningless. Changes in the MCI reflect changing monetary conditions between two points in time. An increase (decrease) in the index indicates that monetary conditions have tightened (eased).

MCI in Practice MCIs figured prominently in monetary policy deliberations at the Bank of Canada and the Reserve Bank of New Zealand in the 1990s. The rationale for constructing these indexes was grounded in the assumption that a central bank can wield considerable control over monetary conditions and that in a small open economy changes in monetary conditions are reflected mainly by changes in the interest rate and the exchange rate, which in turn affect aggregate demand. The level of aggregate demand is the most important factor in determining the rate of inflation. With price stability being an important, if not the overriding, goal of monetary policy, it was incumbent on the central bank to establish desirable monetary conditions that are consistent with price stability. In practice, the central bank prepared a path for the monetary conditions index that was compatible with inflation forecasts. In

the interval between inflation forecasts, the bank monitored the actual MCI to see if it diverged from its desired path or range. In the event that marked differences appeared, the central bank would adjust the interest rate to correct the course of the actual MCI. If evidence emerged for a change in expected inflation, the central bank adjusted the path of the MCI to be compatible with the inflation objective. Given the close attention both central banks paid to the MCI in the day-to-day operation of monetary policy, it seems clear that the MCI served as a short-term operating target.

For other central banks, MCIs have figured less prominently as indicator variables next to the term spread or monetary and credit aggregates. The International Monetary Fund and the Organisation for Economic Co-operation and Development, as well as commercial banks, have also published monetary conditions indexes for various countries. These MCIs serve primarily as information variables: the banks use the actual MCI to assess current monetary conditions and their likely effect on inflation and real economic activity.

Beginning with the new millennium, the prominence of MCIs has waned considerably. Central banks, international organizations, and commercial banks still calculated and monitored MCIs as of 2007, but they played a minor role in monetary policy deliberations. The virtual demise of the MCI is due to several factors, some of which are related to its construction and underlying view of the transmission mechanism. Interpreting observed changes in the index can also be problematic as they depend on one's view of how the economy works.

Role of Exchange Rate An important issue in the construction of a MCI pertains to the determination of the relative weight on the real exchange rate. The two parameters a_1 and a_2 that make up the relative weight are not observable and must therefore be estimated. Whichever relative weight appears in the MCI depends on the specification of the model chosen to estimate aggregate demand. Empirical studies have shown that the range (statistical confidence intervals) for the empirical estimates of the key parameters is excessively wide and often includes zero as a possible estimate. Other issues pertain to the selection of interest rates and exchange rates: real versus nominal, short-term versus long(er)-term interest rates, trade-weighted versus bilateral exchange rates.

The usefulness of a MCI is also hampered by its rather narrow view of the monetary transmission mechanism. In essence, the conception of the original MCI rests on the assumption that the real exchange rate affects aggregate demand and that aggregate demand in turn affects the rate of inflation. Direct effects of the real exchange rate on the rate of inflation through import prices are thus ruled out. Yet in an open economy this direct exchange rate channel plays an important role in determining overall inflation.

The literature explaining the design and use of an MCI warns against mechanically adjusting the interest rate in the wake of a weaker or stronger exchange rate. Careful analysis of the circumstances that gave rise to the change in the exchange rate is warranted before a central bank contemplates a change in the policy instrument. Shocks that originate in the real sector of the economy and lead to changes in the real exchange rate call for a change in monetary conditions, whereas lack of credibility in the conduct of monetary policy or portfolio shocks, that is, shocks in financial sector of the economy, that lead to a weakening of the real exchange rate are to be countered by increases in the interest rate so as to leave monetary conditions unchanged.

It is conceivable that different constellations of the interest rate and the exchange rate could yield the same level of the MCI. In such a situation monetary conditions are the same and the policymaker may be tempted to choose a particular mix of the interest rate and the exchange rate. Yet the outcome for real economic activity and the rate of inflation will most certainly depend on the particular mix chosen, especially if time lags play an important role in the transmission process of monetary policy.

MCIs were still being published around the world as of 2007, because they provide a simple gauge of monetary conditions in an economy. Although MCIs are of some use to financial markets and the general public as an information variable, they play

only a minor role in the implementation of monetary policy at central banks.

See also capital mobility; exchange rate regimes; inflation targeting; interest parity conditions; monetary policy rules; money supply; Mundell-Fleming model; real exchange rate; Swan diagram

FURTHER READING

Deutsche Bundesbank. 1999. "Taylor Interest Rate and Monetary Conditions Index." *Monthly Report* (April): 47 63. A critical assessment of the usefulness of MCIs in the conduct of monetary policy.

Freedman, Charles. 1995. "The Role of Monetary Condi tions and the Monetary Conditions Index in the Con duct of Policy." *Bank of Canada Review* (autumn): 53 59. An early contribution by a central banker that ex plains the rationale for constructing MCIs.

Guender, Alfred V. 2005. "On Optimal Monetary Policy Rules and the Construction of MCIs in the Open Eco nomy." *Open Economies Review* 16: 189 207. This the oretical essay shows that standard MCIs are misspecified if a real exchange rate channel exists in the Phillips curve.

Reserve Bank of New Zealand, Economics Department. 1996. "Summary Indicators of Monetary Conditions." *Reserve Bank (of New Zealand) Bulletin* 59 (3): 223 28. Offers the Reserve Bank of New Zealand's perspective on the usefulness of MCIs.

Stevens, Glenn. 1998. "Pitfalls in the Use of Monetary Conditions Indexes." *Reserve Bank (of Australia) Bulletin* (August): 34 43. A cogent summary of the reasons why central banks should exercise caution in basing the conduct of monetary policy on MCIs.

ALFRED V. GUENDER

■ monetary integration
See common currency

■ monetary policy rules
Monetary policy rules (MPRs) are one of many mechanical processes that guide the implementation of monetary policy. Other processes include a cur- rency board, an exchange rate crawl, or a money growth rate rule. This discussion will be limited to the type of MPR made famous by the Taylor rule (Taylor 1993).

A monetary policy *rule* is an alternative to a *dis- cretionary* policy framework. There are many reasons why rules might possibly be seen as a better option by economists and policymakers. First, rules are more transparent than discretion and, as such, easier to understand and use as the basis for decisions. This, in turn, assists with establishing and maintaining a credible central bank and monetary policy regime. Second, rules are able to convey greater information about the future direction of policy. Third, outcomes under a rules-based policy tend to be better than those from discretionary policy.

Constructing a Monetary Policy Rule An MPR is an algebraic representation of a policy rule that stipulates how an instrument of monetary policy will react to key economic variables and therefore achieve a specific policy objective. There are three main components to the MPR. The first is the policy in- strument. This is the variable that the policymaker adjusts to affect its monetary policy. In the Taylor rule, the MPR is a short-term interest rate.

In the subsequent literature on MPRs, economists have investigated the use of other possible instru- ments of policy. These include the exchange rate, a combination of the interest rate and the exchange rate (a monetary conditions index, or MCI), and even a monetary aggregate such as the monetary base, or M2. Whichever one is used, the instrument is essential for two reasons: it represents the policy lever and therefore must be subject to the policymaker's control, and it must be able to influence economic activity to the extent that it satisfies the policy ob- jectives.

The second component of an MPR is the set of variables to which the policy instrument responds. These variables convey information about the stance of monetary policy and signal how the instrument ought to react so that the policy targets are met. In the basic Taylor rule, the variables to be included are the deviation of inflation from a preannounced target and the deviation of output from a long-run value

designed to represent full employment. Under this specification, the Taylor rule essentially implies that the inflation and output deviations contain sufficient information about the stance of monetary policy to adequately drive interest rate movements.

Although not explicitly stated in the rule, the third component of an MPR is the policy regime that it is supposed to represent. The monetary policy regime most commonly pursued with MPRs is inflation targeting. As such, the boundaries of the MPR reflect how the instrument should react to the right-hand-side variables in a manner that achieves the stated policy.

But how does one link the policy objective to the MPR? The answer is through a statement of central bank objectives usually depicted by a central bank loss function. The loss function is an algebraic representation of the policymaker's objectives and will typically detail the variables that the central bank is interested in targeting. If there are multiple objectives, the weight that the central bank attributes to each objective becomes important. The minimization of the loss function subject to the constraints imposed by a macroeconomic model will yield an MPR.

The MPR emerging from this minimization exercise is known as an *optimal* MPR one that is derived from an explicit policy objective. The Taylor rule itself is an example of a *simple* MPR, one where the rule is specified without reference to explicit objectives. In this instance, the variables captured by the MPR reflect the policy objectives. For example, if the simple MPR contains expressions for output and inflation deviations, then it is assumed that these are the objectives of policy regardless of the model used to depict the macroeconomy. Under an optimal rule, the variables that appear on the right-hand side may not actually be policy objectives they may simply be variables that appear as part of the model used for optimization. Herein lies the difference between simple and optimal MPRs. Optimal MPRs are highly dependent on the model used to derive the rule but are explicitly reflective of the actual policy preferences of the central bank. Simple MPRs are imposed on a macroeconomic model and, as such, do not explicitly reflect central bank policy preferences, but are likely to be robust across many model specifications.

Exchange Rate and Monetary Policy Rules The role the exchange rate ought to play in the adoption of an MPR is a common subject of debate. There are two broad discussions in this area. The first relates to whether the instrument of monetary policy should react in a significant way to the exchange rate when setting policy. The second is whether the exchange rate ought to be used as an instrument, or coinstrument, of policy (see Cavoli and Rajan 2007).

The literature on inflation targeting using MPRs makes it clear that the exchange rate should have no part in the implementation of policy and, if using simple MPRs, should not appear at all in the MPR. The reason for this is that a policy trade-off between the exchange rate and domestic objectives may occur it may be difficult to address both objectives with a single policy instrument. This may well appear to be an appropriate strategy for large, relatively closed economies such as the United States, but the debate is more complicated for open, developing economies. The simple reason for this is that the exchange rate may contain information (about global events or capital flows) that is important in the operation of a MPR that is not contained in other variables (such as output and inflation). This is so even if the exchange rate is not itself an objective of monetary policy. Hence, under these conditions, there is a strong argument for the inclusion of the exchange rate in the rule it helps provide all the necessary information for the attainment of domestic policy.

Once the decision is made to include the exchange rate in the rule, how might one manage the possibility that the exchange rate and inflation may suggest opposing instrument changes? One possibility is to use a partial adjustment. Under this process, the MPR instrument will react to both current and lagged exchange rates. Then, if there is a shock that requires a strong reaction to the current exchange rate but is an inappropriate reaction for inflation, then that reaction can be partially offset next period.

The second issue relates to the possibility of using the exchange rate as an instrument of policy. There are two lines of research here. The first examines a policy rule called a monetary conditions index (MCI). The MCI is an MPR in which the policy instrument is a linear combination of the interest rate and the exchange rate. They are effectively coinstruments of policy. The main premise behind the MCI is that both interest rates and exchange rates indicate monetary conditions and that, for example, a tightening of these conditions can be brought about by an interest rate rise, an increase in the foreign currency value of the domestic currency, or both. As such, it is possible that the variables can move in opposite directions so that monetary conditions are not affected. Therefore, it is the imperative of the policymaker that if a change in one of the variables leads to changes in monetary conditions, the other must be moved to maintain the present policy position and offset monetary changes.

The way that the MCI is typically captured in an MPR is by both the exchange rate and the interest rate appearing on the left side of the rule (see Ball 1999). The economist Laurence Ball (2001), however, believes that this situation is broadly equivalent to an MPR where the interest rate is the main instrument and the exchange rate appears on the right-hand side. Under this version of an MPR, the implicit assumption is that the interest rate is the main instrument of policy and is able to be competently controlled by the policymaker. The exchange rate still offers the policymaker information about the relative tightness of policy and can still guide movements in the instrument and therefore the direction of policy.

The second way that the exchange rate can be used is as a policy instrument in its own right. There is a line of research (particularly in relation to Singapore) where variables such as the nominal bilateral or nominal effective exchange rate (NEER) appear on the left side of an MPR as the sole policy instrument. Underlining this rule is the ability to control the nominal exchange rate (in the conventional manner through foreign reserves). This rule works in precisely the same way as the Taylor-type specification

using the nominal interest rate the right-hand side of the rule would contain expressions for the inflation gap, output gap, possibly lagged exchange rate term, and any other variable deemed to possess information sufficiently important to guide the exchange rate toward the policy objective (Cavoli and Rajan 2006). The overall policy objectives need not be any different from the Taylor-type MPRs; the rule simply reflects the opinion of the policymaker that the exchange rate does a better job at reaching the policy target.

See also exchange rate regimes; inflation targeting; monetary conditions index; money supply; quantity theory of money; time inconsistency problem

FURTHER READING

Ball, L. 1999. "Policy Rules for Open Economies." In *Monetary Policy Rules*, edited by M. Taylor. Chicago: University of Chicago Press, 144 54. Excellent exposition on the construction of open economy MPRs and MCIs.

———. 2001. "Policy Rules and External Shocks." NBER Working Paper 7910. Cambridge, MA: National Bureau of Economic Research. Provides a discussion on open economy MPRs.

Cavoli, T., and R. S. Rajan. 2006. "Managing in the Middle: Characterizing Singapore's Exchange Rate Policy." *Asian Economic Journal* 21 (3): 321 42. Contains an MPR for Singapore where the exchange rate (NEER) is the policy instrument.

———. 2007. "Inflation Targeting Arrangements in Asia: Exploring the Role of the Exchange Rate." Briefing Notes in Economics No. 74 (September October). Presents simple analytics in the construction of closed versus open economy rules and the role of the exchange rate.

Cecchetti, S. 2000. "Making Monetary Policy: Objectives and Rules." *Oxford Review of Economic Policy* 16: 4. Excellent paper on the formulation of closed economy MPRs.

Taylor, J. B. 1993. "Discretion versus Policy Rules in Practice." *Carnegie Rochester Conference Series on Public Policy* 39: 195 214. Seminal paper on the Taylor rule.

———. 2001. "The Role of the Exchange Rate in Monetary Policy Rules." *American Economic Review Papers and*

Proceedings 91: 263 67. Looks at the policy implications of open economy MPRs.

TONY CAVOLI

■ monetary versus fiscal dominance

The terms *monetary dominant* and *fiscal dominant* refer to a central bank's ability to pursue goals of its own choice. Under a monetary dominant regime, the central bank is free to pursue its main goal: stable prices. By contrast, under a fiscal dominant regime, the central bank is not free to fight inflation, since it must print money to fill a fiscal financing gap, now or in the future.

The distinction between monetary and fiscal dominance is of great practical importance for policymakers. For example, central banks have adopted inflation targeting programs in increasing numbers. Inflation targeting, or any other independent monetary regime, is not viable if there are other competing demands on the central bank, such as financing the government's fiscal deficit (when expenditures exceed revenues). In the face of such competing demands, the central bank can limit inflation only in the short run and only with interest rates that are high enough to keep government debt attractive to investors. Upward pressure on interest rates can also be exacerbated if a country suffers capital outflows or adverse shocks to risks as perceived by potential investors in the country.

Economists disagree, however, about how to implement the notions of fiscal or monetary dominance. All agree that markets must be convinced that the central bank is indeed free from other demands. The disagreement lies in exactly how "tight" or "disciplined" fiscal policy must be, currently or prospectively, in order to convey that message. How large must the primary (noninterest) surplus be in order to ensure that the central bank can restrain inflation without excessively high interest rates? The case of Brazil in 2002 3 helps illustrate this point. At that time, Brazil's fledgling inflation targeting program faltered. Even while interest rates were kept high,

the central bank failed to meet its inflation target. Fiscal policy was not overtly "undisciplined." Rather, the government attempted to stabilize its debt when it raised the primary surplus to more than 3 percent of gross domestic product (GDP). Nonetheless, adverse shocks threatened to raise public debt further still, against policy intentions. Fears that the central bank would be forced to buy this extra debt by printing money fiscal dominance thus jeopardized the inflation targeting program (Blanchard 2005).

Explaining Dominance The economists Sargent and Wallace (1981) envisaged a standoff between a central bank and a fiscal authority. They asked what might happen if the central bank signals an unwillingness to satisfy a long-run financing gap. They conclude that, in some cases, even while money is tight and interest rates are high *today*, inflation might still rise *today*.

Their starting point was the government's *intertemporal* budget constraint:

$$B_t/P_t = PV(ps) \qquad (1)$$

where B_t is the nominal value of government claims today (time t), P_t is today's price level, and $PV(ps)$ is the present value of real primary (noninterest) surpluses from today forward. Note that B_t/P_t is real government debt: nominal government debt deflated by the price level. Equation (1) is often called the government's "intertemporal solvency," or "present value," condition. This condition says that, barring default (either explicitly or through a surprise inflation), today's real debt B_t/P_t must be paid for either now or in the future with primary surpluses (ps); the present value of all current and future primary surpluses must be enough to pay off the debt.

Under a monetary dominant regime the fiscal authority meets condition (1) by adjusting the primary surplus (i.e., fiscal balance net of interest payments on existing debt, that is, raising revenues and/or reducing expenditures). If there is an adverse fiscal shock in some period an unanticipated increase in expenditures or fall in tax revenues the fiscal authority should respond by raising taxes or cutting expenditures at some future date. Such an adjust-

ment must offset the initial shock in present value terms. In this sense, primary surpluses are said to be *endogenous*. Since such a policy environment is typically featured in discussions of the Ricardian equivalence hypothesis named after the economist David Ricardo (see Barro 1974), a monetary dominant regime is also called a Ricardian one.

By contrast, under a fiscal dominant (or non-Ricardian) regime, primary surpluses are set exogenously. Even if a shock occurs, the fiscal authority will not adjust its policies. Instead, market participants immediately recognize that the present value of primary surpluses $PV(ps)$ has fallen and is now insufficient to amortize the real value of outstanding debt B_t/P_t. Instead, as equation (1) implies, the real debt B_t/P_t must also fall.

How does this happen? The government might simply announce that it will default on a portion of its debt (or impose a capital levy). A default need not be explicit, however. When market participants themselves recognize that their bonds will not be fully backed by future surpluses, they attempt to reduce their bond holdings. Although any single individual can sell his or her bonds, the market as a whole (in a closed economy) cannot. Instead, when all market participants attempt to sell their debt in exchange for goods, the price of goods will rise just enough to maintain the equality in (1). Sargent and Wallace show how, in such an environment, the inflation rate might rise today, even while money remains tight (see Sargent and Wallace 1981).

The Role of Money and the Fiscal Rule In traditional models, people first exchange their bonds for money and then for goods. In a fiscal dominant regime, money, which is passively supplied by the central bank, bridges the fiscal financing gap. Then, when all agents attempt to purchase goods with their newly printed money, prices rise. Money's role is not essential, however. The economist Woodford (2000) has discussed a case in which paper money is replaced entirely by electronic accounts. According to Woodford, although money itself is not required to pin down the price level, a Ricardian, or monetary dominant, regime is.

Concretely, under such a regime, the primary surplus would be determined according to a rule like:
$$(ps/y)_t = \kappa + \alpha * (b/y)_{t-1} \qquad (2)$$
where ps is the real primary surplus, $b = B/P$ is real debt, and y is real GDP. Such a rule may be tacit; it need not be legally stipulated. Note that, if $\alpha = 0$, for any value of κ, fiscal policy is non-Ricardian. In some countries, $\kappa < 0$; this suggests that the government is giving a rebate to current taxpayers (potentially at the expense of future taxpayers). It can be shown that, for (1) to hold ex ante, primary surpluses and debt must rise together ($\alpha > 0$), but not necessarily on a one-to-one basis ($\alpha < 1$). More precisely, to ensure that (1) holds, α must exceed zero even if only by a small amount.

A rule like (2), in which κ typically equals zero, often appears in the computable general equilibrium real business cycle models that emerged in the 1990s, which are increasingly used by central banks in many countries to guide policy decisions (including inflation targeting programs). Such a rule pins down the price level in the model.

The condition $\alpha > 0$ (κ unrestricted) is a very weak one. For example, if
$$0 < \alpha < [(\rho - \gamma)/(1 + \gamma)] - [\kappa/(b/y)_{t-1}],$$
where $\rho \equiv$ long-run interest rate, $\gamma \equiv$ long-run GDP growth, the debt ratio b/y grows over time. To be sure, the solvency condition (1) is satisfied, but only because primary surpluses ps/y themselves also grow over time passing an ever larger fiscal burden to future taxpayers. A more stringent policy would be one in which the authority attempts to stabilize the debt ratio b/y, specifically $\kappa = 0$ and $\alpha = (\rho - \gamma)/(1 + \gamma)$. This means that the primary surplus is large enough to cover interest payments, minus an adjustment for growth:
$$(ps/y) = [(\rho - \gamma)/(1 + \gamma)]^* (b/y).$$

Such a policy appeals to intergenerational equity: both debt b/y and the primary surplus ps/y stay constant over time. In this way, current and future taxpayers face similar burdens.

Current Practical Approaches for Policymakers In practice, what does it take to reassure market participants that the central bank will not face

pressures to finance deficits? Is it enough that the government has kept itself solvent in the past (i.e., $\alpha > 0$)? Not necessarily. Unless primary surpluses are expected to equal or exceed $[(\rho - \gamma)/(1 + \gamma)]^*(b/y)$, the debt ratio b/y will grow. Such a policy is *unsustainable*: as b/y rises, so too must the primary surplus ps/y. Markets recognize that future fiscal adjustments may be *politically* difficult for the government to achieve.

Even under a debt-stabilizing policy, where primary surpluses equal, on average $(ps/y) = [(\rho - \gamma)/(1 + \gamma)] * (b/y)$, markets may not be convinced that the central bank is free to stabilize prices. This is especially true in developing countries, where external shocks affect debt ratios through their impacts on interest rates, exchange rates, and GDP growth. Often, adverse shocks are not immediately offset by favorable ones. In this case, under a debt-stabilizing rule $(ps/y) = [(\rho - \gamma)/(1 + \gamma)]^*(b/y)$ there is a 50 percent chance that the debt ratio will rise.

This also implies a 50 percent chance that another fiscal adjustment will be needed. Markets may find this probability too high. Governments may wish to reduce that probability. To do so, their primary surpluses must on average *exceed* their debt-stabilizing value. For example, the primary surplus might also be linked to the *volatility* of the fiscal burden (*vol*):

$$(ps/y) = [(\rho - \gamma)/(1 + \gamma)]^*(b/y) + \omega^*(vol),$$

where $\omega > 0$ reflects policymaker preferences. Thus a policy of $\omega^*(vol) > 0$ may be justified as a preemptive cushion that reduces the probability of future adjustments. But, in addition, if $\omega^*(vol) > 0$, public debt will *fall*.

Hence, more recent views have emphasized that, in order to guard against fiscal dominance, governments should aim to *reduce* public debt not just stabilize it. While all governments face a trade-off between costly primary surpluses and the cushion they provide, each will choose a precautionary cushion according to its own circumstances.

In this vein, gauging fiscal dominance (or market perceptions thereof) may require more than simply estimating an equation like (2). For example, some authors have recently developed Monte Carlo techniques (i.e., simulation exercises) that yield estimates of the *probability* that the debt will rise to a certain level during a certain time horizon. And, with an eye toward policy, several papers (see, for example, Tanner and Samake 2008) suggest a method to calculate the primary surplus required to obtain a desired probability. Because such techniques clearly present the trade-off between higher primary surpluses and more security, they should lead to better and more informed policy.

See also debt deflation; discipline; inflation targeting; monetary policy rules; money supply; quantity theory of money; seigniorage; time inconsistency problem

FURTHER READING

Barro, Robert J. 1974. "Are Government Bonds Net Wealth." *Journal of Political Economy* 82 (6) (November December): 1095 1117. A seminal article that spawned a lively debate regarding the Ricardian Equivalence Hypothesis.

———. 1979. "On the Determination of the Public Debt, Part 1." *Journal of Political Economy* 87 (5) (October): 940 71. A seminal article that introduced the notion of fiscal objective functions and "tax smoothing." The article began a series of papers about how to spread the deadweight losses of taxation over time.

Blanchard, Olivier J. 2005. "Fiscal Dominance and Inflation Targeting: Lessons from Brazil." In *Inflation Targeting, Debt, and the Brazilian Experience, 1999 to 2003*, edited by C. Favero and F. Giavazzi. Cambridge, MA: MIT Press, 49 80. A discussion of Brazil's policy dilemmas in recent years.

Celasun, Oya, Xavier Debrun, and Jonathan Ostry. 2006. "Primary Surplus Behavior and Risks to Fiscal Sustainability in Emerging Market Countries: A 'Fan Chart' Approach." IMF Working Paper No. 06/67. Washington, DC: International Monetary Fund. An examination of debt sustainability under uncertainty in emerging markets.

Garcia, Marcio, and Roberto Rigobon. 2005. "A Risk Management Approach to Emerging Market's Sovereign Debt Sustainability with an Application to Brazilian Data." In *Inflation Targeting, Debt, and the Brazilian Experience, 1999 to 2003*, edited by C. Favero and F. Giavazzi. Cambridge, MA: MIT Press, 163 88. This

article presents a "fan chart" approach to examining public debt.

International Monetary Fund. 2003. "Public Debt in Emerging Markets: Is It Too High?" *World Economic Outlook* (September): 113 52. Available at http://www.imf.org/external/pubs/ft/weo/2003/02/pdf/chapter3.pdf. A survey of fiscal policy issues in emerging markets.

Sargent, Thomas, and Neil Wallace. 1981. "Some Unpleasant Monetarist Arithmetic." *Federal Reserve Bank of Minneapolis Quarterly Review* 5 (fall): 15 41. A seminal article that examines the strategic interactions between fiscal and monetary authorities. The piece discusses how inflation could rise even if money was tight if the government budget was not in intertemporal balance.

Tanner, Evan, and Alberto Ramos. 2003. "Fiscal Sustainability and Monetary versus Fiscal Dominance: Evidence from Brazil, 1991 2000." *Applied Economics* 35 (7) (May): 859 73. Features an estimated fiscal reaction function for Brazil.

Tanner, Evan, and Issouf Samake. 2008. "Probabilistic Sustainability of Public Debt: A Vector Autoregression Approach for Brazil, Mexico, and Turkey." *IMF Staff Papers* 55 (1) (February): 149 82. Presents an *objective function* for fiscal policy under uncertainty, corresponding debt sustainability simulations for emerging markets, and a clear menu of options for policymakers.

Woodford, Michael. 2000. "Monetary Policy in a World without Money." *International Finance* 3 (2) (July): 229 60. Presents a model of price level determination in which the money stock does not play a crucial role.

EVAN TANNER

■ money laundering

Money laundering involves the transformation of the proceeds of crime into riches, which can then be enjoyed while disguising their illegal origins. It generally includes three steps: the introduction of the criminal proceeds into the financial system; hiding the proceeds through a variety of transactions and financial vehicles; and investing the laundered proceeds in financial and related assets. These operations often involve international transactions as a means of "layering," or obscuring, the source of the funds. National governments and the international community have worked for many years to combat money laundering, but these efforts were redoubled and expanded to include terrorist financing after the terrorist attacks of September 11, 2001, on the United States.

Measurement of the global volume of money laundering is extremely difficult given that the intrinsic nature of the transactions is intended to disguise the source of funds. Several indirect measurements of the volume of money laundering have been derived, using, for example, estimates of activity in the underground economy and the income generated from criminal activity. The range of these estimates is very broad, from less than one-half to several percent of gross domestic product.

Combating terrorist financing has become an integral part of the international efforts on anti money laundering. Terrorist financing can be defined as the processing of property from any source (perhaps a legitimate one) to finance terrorist activity that has been or will be committed. This activity is thought to use many of the techniques of money laundering and many of the possible countermeasures are similar. Furthermore, several terrorist organizations are known to finance their activities out of the proceeds of crime. Nonetheless, terrorist financing differs from money laundering in several ways that affect public policy. It may be much more difficult to detect terrorist financing than money laundering, since it is mainly directed at future activity: it is possible that the only offense that has been committed when the financing takes place is conspiracy to commit a terrorist act. Also, the amounts of money needed to finance terrorism are widely believed to be relatively small. The September 11 terrorist attack was believed to have required less than $1 million in financing. This compares with the typical volumes of money being laundered by, say, large drug trafficking operations, which in total might approach several hundred billion dollars a year.

International efforts to combat money laundering (and terrorist financing) reflect both a law enforcement strategy to "follow the dirty money" and a

me

concern to protect the global financial system from abuse by criminal elements. When an institution is used unwittingly by criminal elements or terrorists, it risks damage to its reputation. Once the integrity of an institution or financial center is brought into question, its long-term viability is put at risk, with potentially serious economic consequences. If the staff of a financial institution collude with criminal elements to launder funds or channel financing to terrorists, the damage can be much greater. The most serious dangers arise when important financial institutions are controlled by criminals, since these circumstances can compromise the integrity and operations of the whole financial system by corrupting the allocation of resources, misallocating investment, and lowering economic growth.

Combating Money Laundering Money laundering is intrinsically global. If one country or jurisdiction sharpens its focus to prevent money laundering and combat the financing of terrorism (AML/CFT), money-laundering activity will quickly shift to a less-regulated environment. Hence the efforts to combat money laundering and terrorist financing have also been global, centered on the development of international codes and standards that all jurisdictions are expected to follow.

A prerequisite for an effective anti-money-laundering regime is to have in place an adequate general legal framework. Such legislation needs to define and criminalize money laundering and terrorist financing with suitably graduated penalties. It has to cover a wide set of "predicate crimes" that is, the criminal activity that gives rise to the cash or other valuables to be laundered and define the scope of the AML/CFT regime. Commercial banks are generally obligated to be especially vigilant given their role in the payment system. However, as criminals will exploit loopholes, wide sectoral coverage is needed.

The typical measures that financial institutions have to apply include checks on the identity and legitimacy of their clients the "know-your-customer" requirements. Major shareholders and senior managers in financial institutions demonstrate that they are "fit and proper" to hold these positions of control and oversight. Financial institutions must establish systems of identifying and reporting unusual or suspicious transactions. The financial institutions themselves need to train their staff adequately to spot activities that raise a suspicion of money laundering and to have clear processes in place to report back to the authorities. The reporting of unusual transactions and "know-your-customer" rules need to be supported by adequate record keeping. When the authorities, for example, investigate a suspicious transaction, a financial institution must be able to help establish an audit trail.

Typically, financial sector regulators are responsible for supervising the AML/CFT procedures of financial institutions and for checking that their managers and owners meet the "fit and proper" test. Much of this supervision is not separable from other aspects of prudential supervision. For example, a bank supervisor will have to review a commercial bank's internal control procedures to prevent internal fraud or imprudent behavior and, at the same time, can check on whether the bank has in place the means to limit vulnerability to money laundering.

Many countries have also set up specialized agencies called financial intelligence units (FIUs). These agencies investigate, analyze, and pass on to the competent authorities financial and related information concerning suspected proceeds of crime. A key component of an FIU's work is to collaborate with like agencies in other countries in sharing information on suspicious transactions.

A number of organizations are involved in the international efforts on AML/CFT. This includes the Financial Action Task Force (FATF), which sets the global AML/CFT standards; a number of FATF-style regional bodies, such as the Caribbean Financial Action Task Force, which have focused on the development of sound AML/CFT regimes within their regions; as well as the International Monetary Fund, the World Bank, and the United Nations. Together, these bodies provide a network for the global assessment of jurisdictions' compliance with the AML/CFT standards. These assessments identify potential weaknesses in the AML/CFT regimes and develop plans of corrective action. Several of these bodies support the global AML/CFT efforts through train-

ing and technical assistance and through their typologies of money laundering and terrorist financing.

One of the key challenges for any anti-money-laundering regime is that criminal elements are very adept at finding ways to avoid the controls. In an effort to fill possible loopholes in the anti-money-laundering regimes, in 2003 the FATF expanded the AML/CFT standards that countries should comply with to cover a range of new sectors outside of regulated financial institutions, and tightened requirements on existing sectors. The revised standard is comprehensive and demanding, and a number of jurisdictions, especially but not limited to low-income countries, have had difficulty meeting its rigorous requirements. The application of the standard illustrates an ongoing anti-money-laundering issue: how to balance the costs on consumers and industry of tightened regulation with the benefits, when the volume of money laundering has proved difficult to measure.

See also international financial centers; International Monetary Fund (IMF); offshore financial centers

FURTHER READING

Reuter, Peter, and Edwin M. Truman, 2004. *Chasing Dirty Money*. Washington, DC: Institute of International Finance. Provides a comprehensive survey of anti money laundering issues, including a review of effectiveness of the regimes.

R. BARRY JOHNSTON

money supply

Money has been used by civilized societies for millennia, but in its modern macroeconomic context it is identified with "the money supply." Almost all economies have moved beyond commodity money to paper money; however in economies with reliable banks, most money takes the form of bank deposits. To the extent that they are transferable by check, bank deposits satisfy the official definition of money as "a generally acceptable means of payment."

Economists define a country's money supply along a spectrum ranging from high to low liquidity: the ease with which an asset can be used as a generally acceptable means of payment, or at least converted into a means of payment. Various monies can range along a spectrum from coin and paper currency ("cash"), called M0; to M1, which is M0 plus checkable bank deposits; to M2, which is M1 plus savings and other similar short-term but noncheckable deposits; to M3, which includes longer-term deposits; and so on. The precise definitions vary from country to country.

What matters from a macroeconomic point of view is, first, how controllable is the money supply, and second, how much and how quickly it affects output, interest rates, exchange rates, and the overall price level. Governments wish to control their own money supplies in the short run to stimulate output in the face of unemployment, and in the long run to control inflation.

Controllability of Money Supply Controllability of the money supply has long been a matter of debate, but economists generally agree that any central bank with a monopoly on issuance of its domestic currency can control its country's money supply as long as it is not committed to buy and sell foreign currencies at a fixed exchange rate. On the other hand, a central bank that is committed to a fixed exchange rate will be forced to sell or buy unlimited quantities of domestic currency in exchange for foreign currency and will be forced to contract or expand its money supply in the process. The phrase *impossible trinity* refers to the fact that monetary control, fixed exchange rates, and unrestricted international currency flows cannot coexist.

However, controllability is subject to a second constraint. The link between central banks' "instruments" of control and the money supply itself is neither tight nor predictable. One way of conceptualizing this link is via the monetary base, or what is sometimes called "high-powered money": the sum of a country's cash (coin plus paper money, both in circulation and inside banks, or "vault cash") and, more sizably, its central bank's deposits in domestic commercial banks.

Central banks conventionally control money supplies by depositing or withdrawing their deposits with commercial banks, sometimes in exchange for

purchases or sales of government bonds on the open market, and sometimes by other means. Commercial banks then use these "central bank reserves" (defined as central bank deposits plus vault cash) to leverage the creation of deposits   that is, money. They do this by purchasing interest-bearing securities (typically government bonds) and by writing loans, in return for which they issue deposits to sellers of the securities or recipients of the loans. Commercial banks need these reserves at minimum for check-clearing purposes, but also for prudential purposes, against unforeseen cash drains. The central bank often also requires that commercial banks hold a minimum ratio of reserves to deposits.

The problems with trying to control the money supply via the monetary base are threefold. First, the public's desired ratio of cash to deposits is variable and partially unpredictable. When commercial bank deposits expand via an injection of central bank reserves into the banking system, the public will choose to withdraw a fraction of their new bank deposits as cash. This depletes the banks' vault cash and thus their reserves. For prudential and perhaps also regulatory reasons, banks will be forced to reduce their rate of deposit expansion. Hence central banks try to anticipate such cash withdrawals by injecting reserves ex ante, and they also try to reverse unanticipated cash withdrawals by injecting reserves ex post.

Second, the commercial banks' desired ratio of reserves to deposits is also variable and partially unpredictable. Even if, as in many countries, they are required to hold a minimum ratio, they will likely hold some "excess reserves" to guard against unwanted cash withdrawals. How much they choose to hold depends on their opportunity costs: notably, on the yields they earn on alternative assets, such as government treasury bills.

Third, the extent to which banks, after receiving new reserves, choose to expand "demand" or "time" deposits is also variable and unpredictable. Demand deposits (i.e., those on which checks can be written) require more reserves for prudential reasons than do time deposits (i.e., noncheckable) because the likelihood of withdrawals for cash is greater.

In short, the ratio of a country's money supply to its monetary base is variable and partially unpredictable because the public's desired cash-to-deposits ratio, the banks' desired reserves-to-deposits ratio, and the banks' desired mix between demand and time deposits are all unpredictable. Hence the central banks' control of the money supply via the monetary base is imperfect. Of course much of the response of the money supply to the monetary base *is* predictable: it generally rises with income (because the cash-to-deposits ratio falls with income), and it also rises with interest rates (because the excess reserves ratio falls and the time-to-demand-deposits ratio rises). This means that not only is the central banks' job of controlling the money supply made more simple, it also means that money supplies expand and contract endogenously with upswings and downswings in the business cycle, lessening the need for exogenous control by the central bank.

Macroeconomic Implications of Money Supply Growth The second major macroeconomic issue surrounding money supplies is their impact on interest rates, exchange rates, output, and the overall price level. In theory, and subject to many simplifying assumptions, when a country's money supply increases, short-term interest rates will fall in the short run, the exchange rate (in terms of units of domestic currency needed to buy one unit of foreign currency) will rise in the short run, output will rise but perhaps not permanently, and in the long run, overall prices will also rise. The rationale for these effects is that the economy's demand for money depends in the short run negatively on interest rates and in the long run positively on output and overall prices. In the short run, interest rates will fall in order to raise money demand to the new level of money supply, and as a result, the price of foreign currency will rise. In the long run, nominal income   which is the product of output and the price level   will rise. Other things being equal, the increase in nominal income will cease when money demand has increased sufficiently to match the initial rise in money supply. Interest rates and exchange rates will return to their previous levels.

This simple scenario becomes more complex once we consider the dynamics of interest rate and price adjustments between the short run and the long run. In the medium run, the fall in interest rates as well as the increased availability of money for spending will cause macroeconomic demand to increase. This will induce more output   that is, more production of goods and services. The rise in the exchange rate will also increase foreign demand for the country's exports, which will induce a further increase in output. Whether this output increase is temporary or permanent will depend on whether inputs   so-called "factors of production" like labor and capital   are already fully employed. If workers are unemployed and factories are operating at less than full capacity, output can increase permanently. Otherwise, the effect of an increased money supply in the long run is simply to raise prices.

More realistically, we should consider increases in the rate of growth of the money supply, rather than simply its level. In the short run, until factors of production are fully employed, increasing monetary growth increases the rate of increase of output. In the long run, increasing monetary growth increases the rate of increase of prices   that is, it increases the inflation rate   rather than the level of prices per se. Long-run interest rates, which rise with expected inflation, will also rise.

Money Supply in an Open Economy The causes and effects of increased monetary growth are radically modified once we open an economy to inflows and outflows of capital from abroad. If the exchange rate is fixed and foreign currency can readily flow into and out of the country, interest rates will not fall except in the very short run. This is simply a manifestation of the "impossible trinity" mentioned earlier. Any attempt by a central bank to increase monetary growth will result in a small and short-lived decrease in short-run interest rates that will cause foreign currency to flow out until the increase in monetary growth is reversed. Since the central bank is committed to a fixed exchange rate, it will sell enough foreign currency to accommodate the outflow, and in the process it will buy domestic currency,

thereby taking it out of circulation. Interest rates (risk adjusted) will remain unchanged, at "world" levels.

Finally, it is important to understand the effects of the money supply on the exchange rate if it is *not* fixed by the central bank. Briefly, if and when an economy's factors of production are fully employed, a permanent increase in monetary growth will, in the long run, increase the nominal domestic-currency value of foreign currency but leave the "real" value unchanged. A permanent increase in monetary growth is accompanied by a permanent increase in inflation. The country will gradually price itself out of export markets unless it offers an increase in domestic currency per unit of foreign currency equal to the increase in its inflation rate minus any increase in its trading partners' inflation rates. This leaves the "real" exchange rate unchanged.

In short, the rate of growth of a country's money supply is the crucial determinant of its long-run inflation rate. It is also a determinant of short-run interest rates, short-run exchange rates, and short-run rates of output growth. Whether central banks can or should attempt to control short-run interest rates, exchange rates, and output growth by manipulating money supplies is a matter of both theoretical and policy controversy. Indeed, the controversy is moot for central banks that are committed to fixing their exchange rates, since in a world with globalized capital flows they simply have little or no control over their own money supplies.

See also European Central Bank; exchange rate regimes; Federal Reserve Board; impossible trinity; inflation targeting; monetary policy rules; real exchange rate; seigniorage

FURTHER READING
Friedman, Milton. 1968. "The Role of Monetary Policy." *American Economic Review* 58: 1 17. The seminal theoretical statement of relationships between money and real income in the short run, and nominal income and the price level in the long run. The theory linked the short run to the long run via expectations, paving the way for the "rational expectations"

revolution that changed macroeconomics forever in the early 1970s.

Friedman, Milton, and Anna Schwartz. 1960. *A Monetary History of the United States: 1867 1960*. Princeton, NJ: Princeton University Press. This meticulous and now classic study demonstrates convincingly the relationships between the money supply, nominal and real income, and the price level for the United States and, by impli cation, for other countries.

Mishkin, Frederic S. 2004. *The Economics of Money, Banking, and Financial Markets*. 7th ed. Boston: Ad dison Wesley, chap. 3. One of the best textbook treat ments of the institutional arrangements surrounding the creation and control of money in a modern, free market economy.

J A M E S   W.   D E A N

■ monopolistic competition

Most products in the world, even standard com modities such as oil and wheat, are naturally differ entiated because nature is not uniform. But such generic differentiation is quite different from the economic heterogeneity that arises from a deliberate investment in a location, patent, plant, process, or personnel. One must think of the firm as purchasing the differentiation with some sort of investment of fixed resources. This gives rise to monopolistic competition, because other firms can produce a similar or even identical variety at a different location with a high degree of substitution for the original variety. This means that there is free entry but also some degree of monopoly power for each firm. The prototypical example of a differentiated product in international trade is the automobile. Indeed, one can think of the auto industry as monopolistically competitive, with free entry by conglomerate firms such as Toyota, General Motors, Nissan, Mercedes-Benz, and Ford into various segments of the motor vehicle market, such as sport utility vehicles and sport cars. Since an aggregation of customers in every country demands all or most of the varieties offered on the world stage, intraindustry international trade

is the major consequence. In most advanced coun tries, intraindustry trade constitutes more than 50 percent of exports and imports.

Models of Monopolistic Competition The the ory of monopolistic competition was developed in dependently in the 1930s by Robinson (1933) and Chamberlin (1933). This model is strictly partial equilibrium and involves free entry with firms pro ducing output levels that are short of the optimum scale, with the higher costs of production viewed as the cost of variety. To apply the model to interna tional trade, theorists have been quite explicit about tastes, costs, and resources. The most realistic model, offered by Lancaster (1979), assumes heterogeneous consumers with each choosing a preferred variety. This approach, however, is complex because it in volves the distance of each consumer's favored variety from the next best variety to that consumer. Conse quently, the most popular approach has been to as sume a representative consumer who buys some of each variety offered by an industry. This follows the pioneering paper by Dixit and Stiglitz (1977). Dixit and Norman (1980) and Krugman (1979, 1981) applied the Dixit-Stiglitz model to international trade in a way that has now become standard.

In a world of perfect competition, there are gains from trade because a country can import things that would otherwise be produced at home at a higher cost. The gains from trade under monopolistic competition need not be from comparative advan tage, but rather from achieving greater variety and/or lower costs for those differentiated goods. With dif ferentiated products and free entry, the larger market from international trade allows each country to ex ploit economies of scale for some selected products but at the same time give consumers even greater variety from other countries.

The basic idea of the model of monopolistic competition is simple: abstract completely from comparative advantage and make the simplest as sumptions regarding preferences and costs. Thus it is assumed that all products are produced by labor alone, as in the Ricardian trade model, but each va riety has exactly the same production function. Thus

consider the set of products to belong to an industry in which all the varieties are close substitutes. This can easily be captured by a constant-elasticity-of-substitution utility function, $u = \left(\Sigma c_i^a \right)^{1/\alpha}$. With this utility function for each consumer, each variety, regardless of how many, faces the same demand at the same price. As in standard trade models, there is full employment and adjustment costs are ignored in the interests of focusing on the long run. Finally, to capture the features of monopolistic competition, products are always produced under economies of scale and there is free entry and exit of firms. Under these conditions, each firm will be identified with a unique product. Two or more firms could not produce a unique variety unless the average cost curves were U-shaped, which is ruled out by assumption.

Understanding monopolistic competition involves two concepts: economies of scale caused by the fixed costs of setting up a new variety of a particular product, and the elasticity of substitution between the varieties of a certain generic product. The elasticity of substitution between X and Y is simply the percentage increase in the relative demand for good X, defined as X/Y, caused by a 1 percent change in the relative price of good Y, defined by the price ratio p_y/p_x. The higher the elasticity of substitution between two products, the higher the numerical value of elasticity of demand for each of the products. Generally speaking, such a numerical elasticity of demand for good X will be somewhat smaller than the elasticity of substitution, because when the price of good X falls, causing a greater quantity of good X to be demanded, people will buy less of good Y, which increases the elasticity of substitution over the elasticity of demand (because X/Y increases more than X). The more varieties of a single product that exist, of course, the closer the elasticity of demand for any variety will be to its elasticity of substitution with other varieties.

The economics of monopolistic competition is straightforward. Consumers have a love for variety that is reflected in the elasticity of substitution. The higher the elasticity of substitution, the *smaller* the love for variety. With an infinite elasticity of substi-

tution, consumers would be indifferent between the two products: essentially no love for variety. The greater the love for variety, the smaller the elasticity of substitution and, therefore, the smaller the elasticity of demand. But the smaller the elasticity of demand, when firms have control over the price, the higher the markup over marginal cost. Thus the love for variety translates into greater potential profits for introducing a new variety. But this comes into conflict with the fixed cost of introducing the new variety, or the economy of scale. The formula is thus simple: smaller elasticity of demand (greater love for variety) and small fixed costs translate into more varieties of product. Note that the love for variety gives rise to a profit incentive for more variety.

This is where international trade or the size of market comes into play. The larger the market, which is what international trade represents, the less important are fixed costs or economies of scale as an impediment to more varieties of product. Thus international trade results in more varieties being produced, but also more of each variety. When trade opens, the producers of existing varieties find their profits increasing, thus giving an incentive for new firms with different varieties to enter. Consumers are better off because they have more varieties to choose from at lower prices.

Following Dixit and Stiglitz (1977) and Krugman (1979, 1981), the model can be described by just three equations if we assume that all varieties face the same demand and costs: a profit-maximizing equation that is the same for each variety, marginal revenue (MR) equal to marginal cost (MC); a free entry equation, the same price (p) equal to same average costs (AC) for each variety; and a full employment equation, the supply of labor (L) equal to demand, which is simply the number of varieties multiplied by the same labor demand of each variety. Labor is the numéraire, so the wage rate is unity. Let $F =$ fixed costs in terms of the amount of labor required to produce any particular variety, and $c =$ marginal costs in terms of how much labor is required to produce another unit of any variety, which we assume to be constant. If $x =$ the output of each firm,

the $AC = F/x + c$. Each firm will charge a price p that is a markup k over marginal costs. The equilibrium price, p, the output of each firm x, and the number of varieties, n, are then determined by:

$$p = kc \qquad (1)$$

$$p = F/x + c \qquad (2)$$

$$L = n(F + cx) \qquad (3)$$

The markup in equation (1) decreases if there are more varieties. The second equation (2) represents free entry. The last equation (3) is for full employment of the labor resources devoted to the industry. Even though there are three equations, we can represent the model by a diagram with only two variables, the price of each variety and the number of varieties. This is because the last two equations can be used to eliminate x, and to write

$$p = c[(n/L - nF) + 1] \qquad (4)$$

as a substitute for (2) and (3). This equation shows free entry as well as the average cost of production as an increasing function of the number of varieties.

Equations (1) and (4) are shown in figure 1. The curve PP is the price equation (1), which shows that the price of each variety falls as more varieties are produced because individual firms face more competition from substitutes and must lower their prices to maximize profits. The cost equation (4) is CC (ignore the C*C* and C°C° for the moment). It is upward sloping because an increase in the number of varieties, for any labor supply, must reduce the output of each product. But with economies of scale, average costs must then rise so that the free entry price must rise. The equilibrium, a, is the autarky solution for the country in question.

The model is intuitively straightforward. What happens if there is an *increase* in L? This will initially shift all demand curves for the existing n varieties proportionately to the right, without affecting the elasticity of demand at the initial price. Now new firms enter because profits are positive, and this forces all firms to lower p until all profits are again eliminated. Each firm moves down its AC curve and more of each variety is produced, both the old ones and the new ones. With a lower price and more varieties, per capita well-being must increase.

This tells us what happens with international trade. Suppose there are two countries, Home and Foreign. Assume Home and Foreign are identical in all respects but the labor force L, where the foreign L* exceeds home L. In figure 1, PP is the same; but with Foreign being larger than Home, C*C* for the foreign country must lie to the right of CC for the home country since a larger number of varieties is compatible with the same costs. This is clear from equation 4. With free trade and zero costs of transferring goods, the world economy works like one large economy. The cost curve for the world economy is C°C°, which lies to the right of the foreign C*C*. We can see in figure 1 but also by general reasoning that the free trade level of output for each firm will exceed
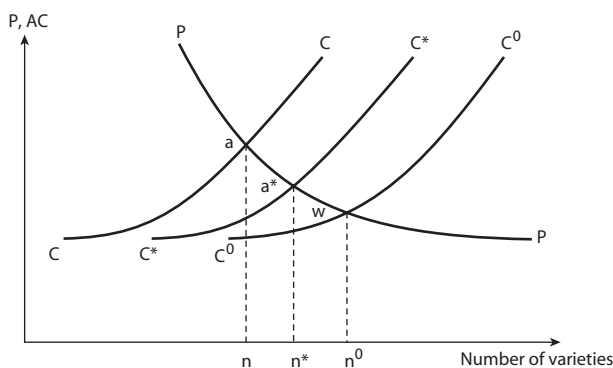
Figure 1

A model of monopolistic competition

the autarkic levels simply because the price of the variety is smaller. But each country will produce a smaller number of varieties because with the same labor force, the only way full employment can be maintained with a larger output of each variety is to reduce the number of varieties produced compared with the number of varieties produced under autarky. Note that with free entry and exit which varieties are produced by which country cannot be determined. To the extent that comparative advantage is not involved, the pattern of trade is indeterminate. When economies of scale are involved, who exports what can be determined by the accident of history.

Now it should be clear that the smaller of the two countries will gain the most from international trade. This is because the price will fall more in that country and the number of varieties will increase. This is clear from the vertical axis of figure 1 and the comparison of the free trade solution at point w with the autarky solution at point a. This is a much stronger theorem in the model of monopolistic competition than in the analogous Ricardian theorem (Ruffin 1988). In the Ricardian model, the result depends on there being only two countries. With more than two countries, the Ricardian result can vanish because the smallest country could have a comparative advantage in between the comparative advantages of the larger countries. But this will never be the case in the present model.

Clearly, this model gives us an explanation of intraindustry trade because in each country all varieties are consumed. All trade is intraindustry in this model because there are no comparative advantages for any industry. Krugman (1981) and Dixit and Norman (1980) show that we can combine intraindustry trade with comparative advantage in a simple way by assuming two such monopolistically competitive industries, but with immobile specific labor assigned to each. Countries can now be defined as having different relative quantities of specific labor, say L_1 and L_2. This allows us to examine the interaction between comparative advantage and intraindustry trade.

Since the price of every variety in this simplified case is the same, we can measure the total output of the industry i as simply the labor supply devoted to that industry divided by the price of the product, p_i, *because if labor is the numéraire, the price of the variety is the amount of labor used in a unit of the product. Thus:*

$$Q_i = L_i / p_i \qquad (5)$$

Since in free trade $p_i = p_i^*$, it follows from equation 1 that

$$Q_i / Q_i^* = L_i / L_i^* \qquad (6)$$

In other words, the relative size of an industry across countries exactly reflects the relative amounts of industry resources across countries. Now if $L_1/L_1^* > L_2/L_2^*$, the home country will on net export commodities in industry 1 and import commodities in industry 2. It should be clear that the greater the difference in factor endowments between countries, the larger the volume of interindustry trade and the smaller the volume of the intraindustry trade.

Thus this simple model can explain why a country such as Japan or Australia would have a much smaller amount of intraindustry trade than the United States or any of the European countries. Japan has a much smaller relative endowment of land and Australia a much larger relative endowment of land than the rest of the world, leading to more interindustry trade because Japan must import more raw materials and Australia must export more primary products.

Intraindustry trade enhances the gains from trade through better exploitation of economies of scale rather than through comparative advantage as trade leads countries to concentrate on a limited number of varieties within any particular industry. This leads to an expansion of world output because of the saving of fixed costs.

As Adam Smith himself observed, specialization within industrial categories may also stimulate innovation more than specialization in primary products such as wheat, oil, and copper. A manufactured good often suggests its own shortcomings or new varieties to the human imagination. Producing a greater variety and number of goods

increases our general knowledge about technology, and greater knowledge implies smaller costs of knowledge accumulation. This has led to the theory of endogenous growth, which augments the model of monopolistic competition with a dynamic structure in which consumers maximize over an infinite horizon and firms innovate to make a profit in such a way that there is still a free entry equilibrium. In this free entry equilibrium, arbitrage profits cannot be made by selling the firm and investing the proceeds. Endogenous growth occurs because every new variety lowers the cost of future new varieties (Grossman and Helpman 1991; Ruffin 1994).

The role of monopolistic competition in international trade is to show how intraindustry trade can arise even in the absence of comparative advantage, and to demonstrate how the gains from trade result from both lower prices and more variety. The model also provides a framework for studying innovation and growth in the world economy.

See also comparative advantage; foreign direct investment under monopolistic competition; gains from trade; New Trade Theory

FURTHER READING

Chamberlin, Edward H. 1933. *The Theory of Monopolistic Competition*. Cambridge, MA: Harvard University Press. A detailed study of both cost and demand conditions in both monopolistic competition and oligopoly.

Dixit, Avinash K., and Victor D. Norman. 1980. *Theory of International Trade*. Cambridge: Cambridge University Press. Applies the theory of monopolistic competition to international trade.

Dixit, Avinash K., and Joseph Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *American Economic Review* 67 (3): 297 308. This paper rescued the theory of monopolistic competition from irrelevance.

Grossman, Gene M., and Elhanan Helpman. 1991. *Innovation and Growth in the World Economy*. Cambridge, MA: MIT Press. Applies various models of monopolistic competition to the problem of explaining some patterns of economic growth in a global economy.

Helpman, Elhanan, and Paul Krugman. 1985. *Foreign Trade and Market Structure*. Cambridge, MA: MIT Press. Considers the impact of oligopoly and monopolistic competition on the volume and gains from trade.

Krugman, Paul R. 1979. "Increasing Returns, Monopolistic Competition, and International Trade." *Journal of International Economics* 9 (4): 469 79. Presents a simple version of the Dixit Stiglitz model of monopolistic competition.

———. 1981. "Intraindustry Specialization and the Gains from Trade." *Journal of Political Economy* 89 (5): 253 66. Applies monopolistic competition to the explanation of intraindustry trade.

Lancaster, Kelvin. 1979. *Variety, Equity, and Efficiency: Product Variety in an Industrial Society*. New York: Columbia University Press, 1979. Applies an innovative theory of product differentiation to the theory of monopolistic competition.

Robinson, Joan. 1933. *The Economics of Imperfect Competition*. London: Macmillan. Presents all of the geometric tools necessary to explain monopoly, monopolistic competition, and price discrimination.

Ruffin, Roy J. 1988. "The Missing Link: The Ricardian Approach to the Factor Endowments Theory of Trade." *American Economic Review* 78 (4): 759 72. Shows how the Ricardian theory of international trade can be used to simplify the Heckscher Ohlin theory and the gains from trade.

———. 1994. "Endogenous Growth and International Trade." *Review of International Economics* 2 (1): 27 39. Summarizes the Grossman Helpman model of innovation and growth in the world economy.

ROY J. RUFFIN

■ moral hazard

See asymmetric information; bail-ins; bailouts

■ most favored nation

See nondiscrimination

■ movement of natural persons

See temporary movement of natural persons

■ multilateral environmental agreements

Since 1933, hundreds of agreements have been negotiated among sovereign nations to protect natural resources and the environment. The United Nations Environment Program (UNEP) (2004) identifies the first multilateral environmental agreement (MEA), dating to 1933, as the Convention Relative to the Preservation of Fauna and Flora in Their Natural State. From 1940 until 1960, UNEP lists an additional 23, including the International Convention for the Regulation of Whaling (1946), the International Convention for the Protection of Birds (1950), the International Plant Protection Convention (1951), and the Antarctic Treaty (1959).

Since 1960, hundreds more have proliferated. Among the most well-known are the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES 1973), the UN Convention on the Law of the Sea (1982), the International Tropical Timber Agreement (1983, amended 1994), the Montreal Protocol on Substances That Deplete the Ozone Layer (1987, 1990, 1992, 1997), the Rio Convention on Biological Diversity (1992), the Kyoto Protocol to the UN Framework Convention on Climate Change (1997), the Basel Protocol on Liability and Compensation for Damage Resulting from Transboundary Movements of Hazardous Substances and Their Disposal (1999), and the Cartagena Protocol on Biosafety to the Convention on Biological Diversity (2000).

Even this small sample of MEAs presents a legal, jurisdictional, and monitoring thicket. Many different international institutions and groups are empowered with implementation and oversight of these agreements, including more than a dozen agencies in addition to UNEP in the UN alone. Apart from the problem of the internal coherence of the many MEAs, throughout the 1990s these agreements were often at odds with other international obligations forbidding restrictions on the free flow of goods and services in international trade, setting up the "trade and environment" debate (Runge 1994, 2001). The World Trade Organization (WTO) and its predecessor, the General Agreement on Tariffs and Trade (GATT), were criticized for undermining environmental obligations under MEAs, as several celebrated cases brought attention to the potential for conflict.

Perhaps the most famous were the pair of "Tuna-Dolphin" disputes in the 1990s, in which environmentalists argued that the GATT articles had been used to overturn a U.S. embargo on tuna harvested by Mexico and other countries using nets that also killed dolphins as by-catch. In imposing the embargo, the United States argued that such fishing methods violated the U.S. Marine Mammal Protection Act (MMPA). The U.S. embargo was overturned by a GATT panel decision in 1994, which concluded that it violated GATT prohibitions on quantitative restrictions to trade and that the MMPA did not qualify for exemptions granted under GATT Article XX(g) "relating to the conservation of exhaustible natural resources."

This infuriated parts of the environmental community, which saw GATT and later the WTO poised to strike down many trade sanctions designed to protect natural resources authorized under both domestic legislation and MEAs.

In 1994, partly in response to this furor, the WTO established a Committee on Trade and Environment (CTE) to respond to these questions. However, as Esty noted, the CTE "has failed to make any policy recommendations of consequence" (Esty 2003). Furthermore, the balance between MEAs and the WTO has been a continuing source of anxiety in environmental circles. In general, analysts have concluded that MEAs lack the coherence, reach, and financial backing of the WTO, leaving them vulnerable to challenge when they attempt to use trade barriers, sanctions, or embargoes as mechanisms of compliance: "Judged in terms of size and teeth," one analyst wrote, "we might regard the WTO as a large tiger and MEAs as a ragged collection of small cats" (Eckersley 2004, 24). This comment is qualified by the observation that there are hundreds of such cats, the herding of which is increasingly difficult.

First, can the tension between MEAs and the WTO be dealt with under existing institutional arrangements? Second, what alternative arrangements might remedy these tensions and provide greater

clarity and coherence to the large number of MEAs? Third, how might these alternatives be approached and successfully implemented in the years ahead?

MEAs and the WTO: Which Way Out? The asymmetrical relationship between the WTO and the multifoliate pattern of MEAs arises from both the structure and content of each set of norms. The structure of the WTO that emerged in the Uruguay Round after nearly 50 years operating under the GATT articles was first developed in the 1947 Havana Charter. In many ways, in 1995 the WTO achieved a vision of an International Trade Organization conceived after World War II as one of three international economic foundations (including the World Bank and International Monetary Fund) to buttress the global economy and resist a return to the beggar-thy-neighbor trade policies of the interwar years. Then, as now, such institutions induced anxiety among those who saw an international trade authority as a threat to national and even intranational sovereignty. In the 1990s, this anxiety extended to national and subnational environmental protection. Whereas the current structure of WTO rules is considered relatively "hard" law in dispute resolution, MEAs are much "softer." And although the contents of MEAs are often tested for consistency with trade rules, trade rules are seldom challenged as inconsistent with MEAs. The result is to discourage the development and implementation of new and existing MEAs and to reinforce their subservience to the trade regime.

There is doubt, given this asymmetry (which many in the trade community consider appropriate), whether a more balanced relationship can be achieved under existing institutions. A number of institutional "fixes" have been proposed. One is to amend GATT Article XX, pertaining to exhaustible natural resources and other exceptions to the rules of nondiscrimination, to better account for MEAs (Nissen 1997). Some precedent for this is reflected in exceptions (article 104) to the North American Free Trade Agreement (NAFTA) for three specific MEAs: CITES, the Montreal Protocol, and the Basel Convention. A less incremental but more direct approach would be to create a new WTO agreement dedicated

specifically to MEAs (Brack and Grey 2003, 36). This could define the general scope of MEAs, the possible use of specific trade measures to enforce or implement their environmental objectives, and criteria for making more nonspecific measures consistent with Article XX. A third proposal, advanced by Switzerland in a submission to the CTE in 2000, would create a clear division of responsibility between the WTO and MEAs, defining their respective trade versus environmental competency. The WTO's responsibility would extend to questions of trade discrimination and disguised protection, while the MEAs' responsibility would be determining environmental objectives and "choosing the means, instruments, mechanisms and measures necessary to achieve these objectives" (CTE 2000, 2). Such an agreement would reduce the asymmetry between trade and environmental measures by asserting the absence of hierarchy and the importance of mutual support and deference.

There is, however, considerable mutual suspicion and derision between the trade and environmental communities resulting from the trade/environment battles of the last two decades. This has led a number of authors to propose the less incremental and more sweeping creation of a World Environment Organization (WEO) (see also Biermann 2000; Esty 2000; Runge 2001).

The WEO Alternative In the early 1990s, transnational environmental policy challenges led to calls for a World Environmental Organization (WEO). By 2000, support extended from the French prime minister Lionel Jospin and president Jacques Chirac to the former WTO director general Renato Ruggiero, the *Economist* magazine, and others. The argument for a WEO arose primarily from those who felt that the WTO system was ill-equipped to respond when trade questions intersected with environmental issues, including MEAs. Apart from trade/environment conflicts, however, the anarchy of the MEA arrangement also suggested a need for better-coordinated multilateral responses to transnational environmental issues even if trade was largely unaffected. In other words, whether or not the WTO could be "greened," international environ-

mental challenges clearly required their own multilateral responses, and these responses required greater coherence and consistency. Just as the GATT/WTO system had evolved out of growing *commercial* interdependence following World War II and had helped to foster a set of rules by which the trade game should be played, so *ecological* interconnections created the need for a more unified set of global environmental rules. The parallelism of trading rules and environmental rules arose from the fact that interdependent states could not cope with commercial *or* environmental challenges solely through unilateral or ad hoc solutions. Moreover, the coexistence of multilateral trade and environmental rules gave rise to the questions of priority and consistency noted earlier.

A prototype for such an overarching institution arose from the environmental side agreement to NAFTA. In late 1993, NAFTA was ratified by the U.S. Congress, together with provisions that would create the North American Commission for Environmental Cooperation (CEC), headquartered in Montreal. By late 2007, the CEC had survived more than a decade of service, although it remained underfunded and often unsupported by the governments of Canada, Mexico, and the United States. Despite its small size, the CEC's mandate extended beyond the trade effects of NAFTA on the environment to include an array of transborder ecological issues. It also provided for a dispute settlement mechanism for environmental questions. From a legal perspective, the CEC derived from an executive agreement, which "steers clear of the normative realm and concerns itself with things institutional, primarily because it is the product of an intergovernmental process between entities that each want to set their own standards" (Johnson and Beaulieu 1996).

The basic design of a WEO modeled on the CEC would be the creation of a Secretariat and a Multilateral Commission on Environment. The Secretariat would be the formal, ministerial-level body of government representatives, meeting periodically to affirm certain policies. The commission would be a policy-oriented group of environmental experts drawn from nongovernmental organizations (NGOs), academia, business, and government. Although the representatives to the WEO Secretariat would, like WTO representatives, be government officials, expert environmental and business involvement, similar to the International Labor Organization (ILO), could occur via the commission. The commission would thus be composed of a standing group of environmental experts and government and business representatives from all member counties. Its meetings would be open to the public and would allow worldwide access to the data and analysis underlying its work. The primary focus of this work would be to propose ways to harmonize national environmental standards and MEAs, while carefully considering the technical issues and problems of this process. This task would resemble in some respects the technical harmonization of food and safety requirements under the Food and Agriculture Organization affiliated Codex Alimentarius. The process would allow for public comments from any group, governmental or nongovernmental, opening the WEO to full public participation and review.

The Years Ahead The CTE's limited terms of reference clearly suggest that the WTO does not care to bring order to MEAs, except insofar as they impinge on trade rules. This substantial institutional gap can be filled by a WEO dedicated to transnational environmental challenges that may or may not have direct linkages to trade, but which have given rise to the crazy-quilt of MEAs.

Until such an overarching structure is articulated, two fundamental questions will challenge MEAs, individually and collectively. First, if MEAs involve explicit trade measures or sanctions, they will be subject to continuing challenges under the WTO. Second, the hundreds of existing MEAs, and those that may be negotiated in the coming decades, cannot be adequately managed without linking them to an overarching set of norms. In the short to intermediate term, waivers to WTO rules have been proposed for some MEAs, renewable over regular time intervals. The presumption that trade ministers should vet each MEA, however, would seem to place them in regular and recurring judgment of

environmental policy measures, a position which they evidently prefer to avoid.

Over the longer term, there appears to be a clear need for the overarching role of a WEO Secretariat to coordinate disparate environmental efforts and MEAs and function as a go-between and buffer relating MEAs or other environmental policies to the WTO system. MEAs (and national environmental programs) designed to minimize trade-distorting effects and advance market-based environmental initiatives such as trading schemes could contribute to this institutional oversight. When obvious conflicts between MEAs and WTO rules arose, a WEO could help to prepare sound arguments in favor of an Article XX exception, based on the necessity of trade measures in an MEA, a waiver if deemed appropriate, or some other GATT-legal "window" allowing an exception for the environmental measure.

An even more significant function might be the consolidation of claims made under myriad environmental agreements into a unified dispute settlement process, in which NGOs and other interested parties could participate. Precisely because an independent entity such as WEO is lacking, a greater temptation exists to use trade measures to enforce environmental obligations. Trade interests may condemn the use of such measures for environmental goals, but in the absence of effective multilateral environmental rules and an overarching entity such as the WEO, environmentalists may claim that they have no recourse. Although it is naïve to imagine that these two realms of policy can be entirely disjoint, the creation of a WEO would assist in separating many issues that do not need to be in conflict. The weaker the perceived ability of environmental groups to influence international policies, however, the greater their incentive to use "linkage" destructively: to threaten the trading system in order to gain environmental concessions.

See also Basel Convention; Convention on Biological Diversity; Convention on International Trade in Endangered Species (CITES); General Agreement on Tariffs and Trade (GATT); Global Environment Facility; trade and the environment; World Trade Organization

FURTHER READING

Biermann, Frank. 2000. "The Case for a World Environmental Organization." *Environment* 42 (9): 23 31. An accessible article laying out the reasons a WEO should be created.

Brack, Duncan, and Kevin Gray. 2003. "Multilateral Environmental Agreements and the WTO." London: The Royal Institute of International Affairs. Sustainable Development Program Report (September). A somewhat technical assessment of the trade issues surrounding MEAs.

Committee on Trade and Environment (CTE). 2000. "The Relationship between the Provisions of the Multilateral Trading System and Multilateral Environmental Agreements (MEAs)." Submission by Switzerland (June 8) (WT/CTE/W/139). A trade negotiator's note on trade and MEAs.

Eckersley, Robyn. 2004. "The Big Chill: The WTO and Multilateral Environmental Agreements." *Global Environmental Politics* 4 (2): 24 50. A political scientist's analysis of MEAs and conflicts with trade rules.

Esty, Daniel C. 2000. "The Value of Creating a Global Environmental Organization." *Environment Matters* (annual review). Washington, DC: World Bank. A former trade negotiator argues for a world environmental body.

———. 2003. "Introduction: Achieving Harmony in Trade and Environment." In *GETSView* (online journal) 2 (2). Downloadable from http://www.GETS.org/pages/harmony/. An introduction to a collection of papers on trade environment interactions.

Johnson, Pierre Marc, and André Beaulieu. 1996. *The Environment and NAFTA: Understanding and Implementing the New Continental Law*. Washington, DC: Island Press. The most complete analysis of the environmental side agreement to NAFTA in print.

Nissen, J. L. 1997. "Achieving a Balance between Trade and the Environment: The Need to Amend the WTO/GATT to Include Multilateral Environmental Agreements." *Law and Policy in International Business* 28 (3): 901 28. A legal analysis of the WTO/MEA relationships.

Runge, C. Ford (with François Ortalo Magné and Philip Van de Kamp). 1994. *Freer Trade, Protected Environment: Balancing Trade Liberalization and Environmental*

Interests. New York: Council on Foreign Relations Press. A monograph resulting from the Council on Foreign Relations study group on trade and environment.

———. 2001. "A Global Environment Organization (GEO) and the World Trading System." *Journal of World Trade* 35 (4): 399–426. The author's arguments in favor of and against a global environment organization.

United Nations Environment Program (UNEP). 2004. "Chronological List of Environmental Agreements." Division of Policy Development and Law. A comprehensive listing of all MEAs to 2004.

C. FORD RUNGE

■ **multilateral trade negotiations**

Broadly defined, a multilateral trade negotiation is any trade negotiation occurring between more than two countries. However, since the advent of the General Agreement on Tariffs and Trade (GATT) in 1947, the term has come to refer specifically to the periodic negotiating "rounds" (meetings) conducted under the auspices of the GATT and successor, the World Trade Organization (WTO). There have been nine such rounds since 1947, with the most recent—the Doha Round—not yet completed as of 2007. The purpose of these rounds is to reduce barriers to international trade and establish or modify rules of conduct for trade policy. What distinguishes them as multilateral is that participation is open to all member countries and the results generally apply to all members on a nondiscriminatory basis. As each successive round builds on the results of the one before, the GATT/WTO system evolves, as both the product of multilateral trade negotiations and the framework in which the negotiations are conducted.

Basic Concepts and Terminology Throughout history, countries have sought to obtain access to foreign markets for the goods and services they export, while protecting their own domestic markets from foreign competition. This behavior is reflected in the terminology of trade negotiations. If a country lowers one of its tariffs or otherwise opens its market to imports from another country, it is said to have granted a "concession" to that country. Trade negotiators view receiving a concession as beneficial and granting a concession as costly, though the magnitude of the benefits and costs may vary from sector to sector. The mutual exchange of concessions of roughly equal value is known as "reciprocity."

Both economic theory and ample empirical evidence have shown that, in general, concessions produce welfare benefits to the affected exporting countries that outweigh whatever welfare costs are experienced by the importing country granting the concession, provided the concessions are nondiscriminatory, that is, provided the concession is granted to all exporters. This suggests that if all countries exchange concessions on a reciprocal and nondiscriminatory basis, then all stand to benefit. This insight is at the heart of multilateral trade negotiations. Indeed, the stated purpose of the GATT is to raise incomes, employment, and growth "by entering into reciprocal and mutually advantageous arrangements directed to the substantial reduction of tariffs and other barriers to trade and to the elimination of discriminatory treatment in international commerce" (Preamble of the General Agreement on Tariffs and Trade, 1947).

Pre-GATT Trade Negotiations Prior to the GATT, countries negotiated trade agreements on a bilateral basis. Two countries interested in expanding trade with each other would meet and negotiate reductions in trade barriers, usually tariffs, on a more or less reciprocal basis. Early trade agreements negotiated in this way were relatively few in number, short in duration, and usually discriminatory (i.e., concessions applied only bilaterally). This changed substantially in the mid-19th century, however. A surge in global trade, fueled by the industrial revolution and technological improvements in transportation and communications, coincided with changes in the political and intellectual landscape in favor of freer trade. Many countries responded by opening their markets, both unilaterally and through bilateral agreements.

The most important bilateral agreement of the period was the Cobden-Chevalier treaty of 1860 between the United Kingdom and France, which cut

tariffs between two of the largest trading nations of the time. It also included the most-favored-nation (MFN) clause, under which both countries committed to giving each other the best treatment they offered to any other country. This had the effect of guaranteeing that neither country would discriminate against the other in any subsequent agreements they might negotiate with other countries. A wave of bilateral MFN agreements followed, as countries rushed to avoid being the victims of discrimination, so that by 1870, almost all major traders were involved in a web of nondiscriminatory bilateral agreements.

The successful bilateral trade regime of the 19th century was suspended during World War I and suffered a series of setbacks until the end of the World War II. The most notable setback was a global trade war touched off by the U.S. Smoot-Hawley Tariff Act of 1930. Four years later the United States attempted to reverse the slide toward protectionism with the passage of the Reciprocal Trade Agreements Act (RTAA), which gave the president wide latitude to negotiate bilateral reciprocal trade agreements on an MFN basis. Although this effort to revive the bilateral system was cut short by World War II, many of the concepts, rules, and procedures of the RTAA were subsequently incorporated into the GATT after the war.

Negotiations in the GATT Era In 1946, the United Nations, responding to a proposal from the United States, established a preparatory committee to draft a charter for an International Trade Organization (ITO). The ITO was meant be a multilateral trade organization with open membership, founded on the principles of reciprocity and nondiscrimination, to serve as the trade counterpart to the organizations set up by the Bretton Woods conference of 1944 the International Monetary Fund and the World Bank. At its first meeting in London, the preparatory committee established procedures for conducting multilateral trade negotiations, which were put to use at a meeting held in Geneva in 1947 involving 23 countries: Australia, Belgium, Brazil, Burma, Canada, Ceylon, Chile, the Republic of China, Cuba, the Czechoslovak Republic, France, India, Lebanon, Luxembourg, Netherlands, New

Zealand, Norway, Pakistan, Southern Rhodesia, Syria, South Africa, the United Kingdom, and the United States. The countries exchanged concessions in the form of tariff "bindings" maximum limits on tariff levels covering thousands of manufactured products. They also agreed that all signatories should grant one another MFN status, and they drafted a set of general rules on the use of other trade-related policies, such as subsidies and quantitative restrictions, to ensure that countries did not circumvent their tariff obligations. The concessions and rules together constituted the General Agreement on Tariffs and Trade (GATT), and the Geneva meeting became known as the first GATT round. The GATT was supposed to be incorporated into the ITO Charter. Although the charter was finally completed in Havana in 1948, it was never ratified, however. Thus, the GATT emerged as the basis of the multilateral trading system.

Subsequent rounds of multilateral trade negotiations were held in Annecy, France (1949), Torquay, United Kingdom (1950), and Geneva (1956). After this, rounds were given names: the Dillon Round (1960 61) was named after the U.S. undersecretary of state C. Douglas Dillon, the Kennedy Round (1963 67) after the U.S. president, and the Tokyo (1973 79) and Uruguay (1986 94) Rounds after the places where they were launched. The Doha Development Round, so named because issues of developing countries were supposed to be high on the agenda, was launched in Doha, Qatar, in 2001. A summary of these rounds is found in table 1.

Subjects and Modalities At the beginning of each round, the participating countries convene to decide on the subjects to be negotiated and the "modalities" the rules and procedures for conducting the negotiations. In the first five GATT rounds, the primary subjects were tariff concessions on manufactured products and the accession of new members. New members are required to negotiate tariff concessions and other terms with existing members prior to their entry. In the Kennedy Round, the scope of the negotiations was broadened to include concessions on nontariff barriers to trade in manufactures. The Tokyo Round continued this

Table 1
GATT/WTO rounds, 1947 2007

Name of round or meeting	Period	Number of parties	Subjects and modalities	Results
Geneva	1947	23	Tariffs: item by item request offer negotiations	Concessions covering 15,000 tariff lines
Annecy	1949	33	Tariffs: item by item request offer negotiations	5,000 tariff concessions; 9 accessions
Torquay	1950	34	Tariffs: item by item request offer negotiations	8,700 tariff concessions; 4 accessions
Geneva	1956	22	Tariffs: item by item request offer negotiations	Modest reductions
Dillon Round	1960 61	45	Tariffs: item by item request offer negotiations.	4,400 tariff concessions.
Kennedy Round	1963 67	48	Tariffs: formula approach (linear cut) and item by item negotiations. Nontariff measures: anti dumping, customs valuation	Average tariffs reduced by 35%; 33,000 tariff lines bound; agreements on customs valuation and anti dumping
Tokyo Round	1973 79	99	Tariffs: formula approach with exceptions. Nontariff measures: anti dumping, customs valuation, subsidies and countervail, government procurement, import licensing, product standards safeguards, special and differential treatment of developing countries	Average tariffs reduced by one third to 6 percent for OECD manufactures; voluntary codes of conduct agreed for all nontariff issues except safeguards
Uruguay Round	1986 94	117	Tariffs: formula approach and item by item request offer negotiations. Nontariff measures: all Tokyo issues, plus services, intellectual property, preshipment inspection, rules of origin, trade related investment measures, dispute settlement, transparency and surveillance of trade policies	Average tariffs again reduced by one third; agriculture and textiles and clothing subjected to rules; creation of WTO; new agreements on services and TRIPs; majority of Tokyo Round codes extended to all WTO members
Doha Round	2001	150	Tariffs: formula approach and item by item request offer negotiations. Nontariff measures: trade facilitation, rules, services, environment	

Source: World Trade Organization, *World Trade Report 2007.*

trend and added voluntary codes of conduct on various trade-related policies. The Uruguay Round expanded into the areas of agriculture, services, intellectual property, and trade-related investment measures.

Elaborate modalities have been developed for the negotiation of concessions on market access. The first five rounds used the procedure known as the item-by-item, request-and-offer method. Under this method, each participating country simultaneously submits a "request list" to each of its trading partners. A request list is a list of concessions (products and tariff cuts) that the country requests from its trading partners. Each trading partner responds with an offer list, a list of concessions that the country is willing to make. Each participating country has full flexibility on granting concessions on individual products and is free to refrain from concessions on any item. Moreover, countries may request concessions only on products for which they are among the largest suppliers to the countries from which the concessions are asked. This is known as the "principal supplier" rule. The reason for this rule is that all concessions in GATT negotiations must be made on an MFN basis, which means that they must be extended to all GATT members. If a small supplier were to request a concession, then to grant that concession would imply that most of the benefit, in terms of improved market access, would accrue to countries (the large suppliers) that did not request the concession and therefore would be unlikely to "pay" for it with reciprocal concessions. Countries that receive concessions via MFN without paying for them are free riders. The principal supplier rule is a method for containing the free-rider problem.

The Kennedy Round brought two significant changes in modalities. First, it was agreed that negotiations would center on across-the-board linear tariff cuts. Countries were expected to cut tariffs on all manufactured products by an equal percentage (50 percent was the agreed number) but could submit lists of exceptions. These lists are referred to as "negative lists" (lists of tariffs *not* to be reduced, or reduced by less than 50 percent), as opposed to the "positive lists" (list of tariffs to be reduced) in the

item-by-item approach. The second change was the adoption of special and differential treatment for developing countries. In particular, it was decided that less-developed countries should not be expected to reciprocate tariff reductions made by developed countries. These changes in modalities reflected a growing concern about the unwieldiness of the item-by-item approach as the number of GATT members grew. They also reflected concern among developing countries that trade liberalization would interfere with their development policies and that negotiations based on the principal-supplier rule had put them at a disadvantage.

The Tokyo Round operated with a similar procedure as the Kennedy Round, except that instead of a linear tariff cut, countries used a more complicated formula designed to achieve larger reductions on tariffs that were initially high. This became known as the "formula" approach. The Uruguay Round involved a mixture of the formula approach and the item-by-item approach. Some countries, led by the European Union, used the formula approach with negative lists, while the United States and other countries used the item-by-item approach, submitting positive lists.

Negotiations are normally carried out by representatives from each individual country. However, in some cases, countries join coalitions and negotiate as a group. The most common form of this is the customs union. A customs union is a group of countries with free trade among them and a common external tariff. Examples include the European Union, the South American customs union Mercosur, and many others. Typically they send a single representative to negotiate changes in the common external tariff. Other coalitions will typically form around a common interest in a specific topic. For example, agricultural exporters have created coalitions to make common offers on agricultural tariffs and subsidies. Small countries often create coalitions to gain greater weight in negotiations.

Results Of the first five GATT Rounds, the first was by far the most productive in terms of tariff concessions. It involved on the order of 45,000 concessions on 15,000 tariff lines, covering about

$10 billion worth of trade. Such extensive tariff reductions were not repeated until the Kennedy Round, at which point the GATT had more than twice the number of members. The lack of productivity of the negotiations during the intervening rounds, however, does not mean that trade liberalization was halted. In fact, tariff cuts agreed to by Europe and Japan in 1947 were rendered largely ineffective by quantitative restrictions and exchange controls in those countries, until the late 1950s. As these measures were eliminated on a unilateral basis in the late 1950s and early 1960s, the international trade of these countries soared.

The linear approach to tariff reductions used in the Kennedy Round proved effective, though tariffs were ultimately reduced by an average of only 35 percent, instead of the 50 percent target, and many countries opted out of the linear reduction entirely. This round also produced agreements on customs valuation and anti-dumping. The Tokyo Round achieved similar tariff reductions and added voluntary codes of conduct on subsidies, technical barriers to trade, import licensing, government procurement, customs valuation, anti-dumping, meat and dairy products, and civil aircraft. The Uruguay Round cut tariffs again by a third, created a new organization, the WTO, with a revised organizational structure, and strengthened the dispute settlement system.

The creation of the WTO did not fundamentally change the nature of multilateral trade negotiations; however, it did create additional opportunities for countries to discuss trade policy on an ongoing basis. An important innovation is the Ministerial Conference, which meets every two years. This conference is attended by the top trade representative of each country and is intended to provide direction and political impetus for moving trade negotiations forward. The Doha Development Round was launched at the Doha Ministerial Conference in 2001.

It is fair to conclude that multilateral trade negotiations have played, and continue to play, a key role in shaping the modern world trading system. They have established the general rules and terms that now govern the trade of more than 150 countries and have brought about dramatic, if uneven, reductions in tariffs and nontariff barriers since 1947. Whether they will continue to work as the primary mechanism for trade liberalization in the future is unknown. Increases in membership and greater complexity of the issues under negotiation have made recent multilateral negotiations slow and difficult. Preferential trade agreements, such as customs unions and free trade areas, negotiated bilaterally or among small groups of countries, have been on the rise since the mid-1990s. They offer greater speed and focus than negotiations involving the whole WTO. Whatever the ultimate effect of preferential arrangements, however, multilateral trade negotiations will continue to be important, because they are the only option for addressing trade issues of global consequence.

See also agricultural trade negotiations; General Agreement on Tariffs and Trade (GATT); multilateralism; nondiscrimination; tariff-cutting formulas; World Trade Organization

FURTHER READING

Bagwell, Kyle, and Robert W. Staiger. 2002. *The Economics of the World Trading System*. Cambridge, MA: MIT Press. Suitable for advanced students of economics, this book provides a theoretical framework for understanding the economics of trade negotiations, agreements, and WTO rules.

Hoda, Anwarul. 2001. *Tariff Negotiations and Renegotiations under the GATT and the WTO: Procedures and Practices*. Cambridge: Cambridge University Press. An excellent source for the details, large and small, of multilateral trade negotiations.

Hoekman, Bernard, and Michel Kostecki. 2001. *The Political Economy of the World Trading System*. 2d ed. New York: Oxford University Press. Offers a broad introduction to the WTO, along with an accessible treatment of the economics and politics of multilateral trading system.

Jackson, John H. 1997. *The World Trading System: Law and Policy of International Economic Relations*. Cambridge, MA: MIT Press. A classic, this book gives a broad and insightful introduction to the history and institutions of the GATT/WTO, organized by policy type.

World Trade Organization (WTO). 2007. *World Trade Report 2007: Sixty Years of the Multilateral Trading System, Achievements and Challenges.* Geneva: WTO. Written by a team of WTO experts, this book offers thorough treatment of the history of the GATT and WTO.

RODNEY LUDEMA

■ **multilateralism**

With a membership of more than 150 nations in the World Trade Organization (WTO) and a broad mandate covering trade in goods and services, the multilateral trade system has come a long way since its inception in the early postwar years as the General Agreement on Tariffs and Trade (GATT) a provisional treaty that was signed by merely 23 founding member countries (contracting parties).

Under the GATT, member countries agreed to adhere to specific rules and principles relating to goods trade and to participate in periodic rounds of multilateral trade negotiations to lower trade barriers. The Uruguay Round of multilateral trade negotiations (1986 94) created the WTO subsuming the GATT under it and adding two new agreements for the WTO to oversee: the General Agreements on Trade in Services and the Agreement on Trade-Related Aspects of Intellectual Property Rights. Although the WTO is a complex and multifaceted institution, its core purposes are to serve as a forum in which countries reach *agreement* on rules governing their trade policies and on reductions in their trade barriers and to *enforce* these agreements.

It may seem surprising that the theoretical literature on international trade relations does not unconditionally prescribe a multilateral institution to mediate trade relations among nations. Specifically, it is well known that in perfectly competitive contexts, a "small" country maximizes its national welfare by *unilaterally* choosing free trade. This benchmark setting offers no reasons for trade negotiations between countries. When a country is "large" enough to influence prices in international markets, however,

a rationale for trade negotiations emerges. Although the exercise of market power by a large country through the imposition of trade restrictions leads to an improvement in its terms of trade (the relative price of its exportables to its importables) and welfare, global welfare declines. Large countries may therefore find themselves in a "prisoner's dilemma": individually rational behavior leads them to impose trade barriers on one another, even though they would all be better off (relative to this outcome with mutual barriers) by lowering their trade barriers instead (Johnson 1953 54; Bagwell and Staiger 1999). Clearly, negotiations may help to move these countries to the superior outcome with lower trade barriers. It may be noted that domestic political factors may prevent even a "small" country from achieving free trade. Here, too, international negotiations involving reciprocal trade liberalization by partner countries may induce countries jointly to liberalize trade by suitably altering the factors influencing domestic politics and decision making.

The WTO system provides mechanisms in the form of explicit and implicit rules, norms, and principles to increase the incentives for agreements to be reached and for agreements to be enforced. Primary among these are *reciprocity* and *nondiscrimination*. Although *reciprocity* has not been formally defined in the WTO, in practice it calls for a balance of "concessions" to be made by the parties to an agreement. Since the initial trade barriers may be suboptimal to begin with, lowering them may improve welfare, and thus this need not be regarded as a "concession" in the first instance. Nondiscrimination, formally explicated in the GATT (Article I), requires that no member country be discriminated against specifically that imported goods need to be treated equally regardless of country of origin.

In the multilateral system, reciprocity and nondiscrimination alter the incentives for agreement to be reached and for agreements to be enforced. Where trade barriers come about due to countries' exercise of market power, unilateral liberalization may reduce welfare by worsening the liberalizing country's terms of trade. Reciprocal trade liberalization, on the other hand, will allow countries to maintain their terms of

trade while achieving enhanced production and consumption efficiency. When tariffs are determined by domestic political factors, such as the political strength of the import-competing sector, reciprocal liberalization, by improving outcomes for domestic exportable sectors, enhances the prospects for trade liberalization. Note that nondiscrimination can add significant leverage to reciprocity. By ensuring that one country's liberalization extends to all partners, nondiscrimination creates a wider set of reciprocal liberalization obligations. This, in turn, again due to nondiscrimination, creates an even wider set of reciprocal liberalization obligations. Nondiscrimination coupled with reciprocity may thereby create a virtuous circle of liberalization. On the other hand, some argue that nondiscrimination creates "free-rider" problems within the system, since countries not engaging in liberalization of their own nevertheless benefit from the liberalization of others. In practice, however, the scope for free riding is minimized by exercising selectivity and exclusion of goods chosen for liberalization (that is, a country may liberalize trade only in those specific goods that are relevant to a reciprocating partner country and exclude those goods that may be of value to other nonliberalizing countries).

Enforcement of WTO agreements is achieved through a well-specified dispute settlement mechanism that involves a series of steps progressing through consultation between members, adjudication by WTO panels, and adoption by its dispute settlement body. Thus transgressions of WTO agreements require that the affected parties be offered compensation by the transgressor or that they be allowed reciprocally to withdraw "equivalent concessions" (compensation may be preferred, as raising trade barriers in retaliation may be costly for the retaliating country as well).

Nondiscrimination implies that more than one country will be affected by transgression of WTO agreements the withdrawal of equivalent concessions by all of these affected parties increases the costs of violating WTO agreements. Multilateral punishments may also be valuable in sustaining the system and "leveling the playing field" when there are trans-

gressions of WTO agreements on a bilateral basis (discussed later) and when countries are asymmetric in their economic power or vulnerability (Maggi 1999). Thus reciprocity and nondiscrimination acting together play an important role in the enforcement of agreements as well.

WTO agreements provide a number of provisions that allow for temporary, and sometimes even permanent, suspension of previously undertaken obligations. To protect industries against a sudden surge of imports, the GATT permits countries to take "safeguard" actions that are temporary, nondiscriminatory, and transparent (Article XIX). The GATT also allows for anti-dumping duties to be imposed on imports that are being sold at prices lower than costs or lower than prices in the exporter's home market (Article VI). Countervailing duties may be imposed on imports from suppliers who have been "unfairly" subsidized by their governments (Article VI). In contrast to "safeguard" actions, which must, at least in principle, be applied on a nondiscriminatory basis, anti-dumping and countervailing duties are generally country specific. The multilateral retaliation that supports enforcement of agreements in the multilateral trade system is clearly weakened in these cases.

In a significant derogation of the GATT's nondiscrimination principle, member countries are permitted to enter into preferential trade agreements (PTAs) in the form of *free trade areas* and *customs unions* (Article XXIV). Although PTAs are subject to the requirement that trade restrictions imposed by the members of the preferential agreement on nonmember countries will not be more restrictive than before the preferential agreement and a further restriction that the member countries eliminate barriers to "substantially" all trade among themselves, violations of these restrictions abound with the WTO showing limited enforcement ability.

PTAs have rapidly proliferated in recent years, as have uses of nontraditional (and bilaterally implemented) forms of protection such as anti-dumping and countervailing duties. Reining in these significant challenges to multilateralism will require the multilateral trade system to strengthen its

enforcement mechanisms and to revitalize interest in its nondiscriminatory liberalization mechanisms.

See also Doha Round; General Agreement on Tariffs and Trade (GATT); multilateral environmental agreements; multilateral trade negotiations; nondiscrimination; regionalism; Uruguay Round; World Trade Organization

FURTHER READING

Bagwell, K., and R. Staiger. 1999. "An Economic Theory of GATT." *American Economic Review* 89 (1): 215 48. A theoretical rationalization of the principles of reciprocity and nondiscrimination in the international trade system.

Hoekman, B. 1993. "Multilateral Trade Negotiations and the Coordination of Commercial Policies." In *The Multilateral Trading System: Analysis and Options for Change*, edited by R. Stern. University of Michigan Press, Ann Arbor, 29 62. A comprehensive discussion of the rationale for multilateralism and the relevant char acteristics of the GATT/WTO system.

Johnson, H. G. 1953 54. "Optimum Tariffs and Retalia tion." *Review of Economic Studies* 1 (2): 142 53. A sem inal theoretical analysis of self interested behavior of countries in setting trade policy and inefficiency of the resulting policy outcomes.

Maggi, G. 1999. "The Role of Multilateral Institutions in International Trade Cooperation." *American Economic Review* 89 (1): 190 214. A theoretical analysis of the role of multilateral enforcement in sustaining trade cooper ation.

PRAVIN KRISHNA

■ multinational enterprises

The multinational enterprise (MNE), also referred to as the multinational corporation, the international corporation, the global firm, or the transnational corporation, is a multiplant firm that controls and co ordinates operations in at least two countries (Caves 2007). Although there is a tendency to associate the control and coordination of an MNE's inter national operations with majority-owned foreign affiliates (undertaken through foreign direct invest ment, or FDI), both control and coordination may be achieved through a minority ownership and in some cases through nonequity means.

Historically, FDI and MNE activity have been synonymous. International and national agencies that maintain and collect data on MNE activity generally define the MNE as a firm that *owns* and controls value-adding operations in more than one country (Dunning 1993). Although FDI is one of the main modes by which MNEs engage in cross-border value-adding activities, the MNE may also control and engage in value-adding activities through nonequity means, such as through co operative agreements and outsourcing, sometimes without de jure ownership of the productive assets but de facto controlling the operations of the nonlegally affiliated operation. Therefore, the use of the term *MNE* as a synonym for FDI is in creasingly inaccurate (Wilkins 2001).

The MNE has traditionally also been regarded as having a distinct "home country" where its head quarters are located and which acts as the command center, providing primary strategic direction for its affiliates in various "host countries," as well as pro viding the primary or core assets on which the affil iates base their operations. Indeed, the home country establishment is referred to as the "parent firm," implying that the home country is at the top of a hierarchy of affiliates or subsidiaries. Furthermore, residents of the home country are assumed to own (and therefore control) the international operations of the MNE. Although most MNEs are still orga nized along these lines, there are a growing number of firms in which ownership and control are spread across several countries, as well as several cases where an MNE may locate its headquarters in a country other than its home country. Furthermore, a number of MNEs have multiple headquarters located in different countries, with strategic control distributed along geographic, product, or functional lines.

The nature, structure, and organization of the MNE has changed considerably since it has become an object of specialist study, and this change has been exacerbated by development associated with the globalization of products, markets, and services, as well as political and economic developments associ-

ated with shifting political hegemonies and economic liberalization (Narula 2003). Thus a more accurate and current definition of an MNE is: a firm that engages in international business activities, has affiliates in more than one country, and whose operations and activities in different locations are actively coordinated by one or more headquarters organizations. Emphasis is placed on the presence of interdependencies between the various economic units in different locations, and their active coordination and control across borders, and not on the ownership structure (Narula 2003).

Significance of MNEs United Nations Conference on Trade and Development (UNCTAD) data indicate that in 2004 there were an estimated 77,135 MNEs worldwide, with 773,019 affiliates of various types (UNCTAD 2006). Although MNEs are often regarded as large, global firms, the majority of MNEs are small and medium-sized enterprises, with relatively small international operations, often concentrated in their home region.

Despite the importance given to MNEs, they do not account for a dominant or even a major share of the world's economic activity. In terms of employment, the world's largest nonfinancial MNEs employed 14.9 million people in 2004. This number represented only 6 percent of total global employment. Despite the relatively small role of MNEs on an aggregate level, the situation differs substantially across countries. MNEs account for less than 5 percent of the total employment in countries such as Japan and Indonesia, but this figure rises to well over 40 percent in countries such as Malaysia, Argentina, and Ireland. The significance of MNEs in terms of sales, value added, and ownership of assets is of similar magnitude. In the United States, value added, exports, imports, and research and development (R&D) by foreign MNEs in 2002 as a percentage of total activity by all private businesses were 5.7, 19.8, 27.9, and 14.4 percent, respectively.

Two characteristics of the operations of MNEs stand out. First, larger MNEs have been found to be concentrated in the more "dynamic" sectors of the economy. Thus even though they play a relatively small role in most economies in terms of the level of total employment, MNEs often play a disproportionately large role in two different types of industrial sectors. They tend to be most active in dynamic sectors typified by high growth rates and the use of new and emerging technologies (e.g., electronics, communication equipment, and industrial machinery), and also in mature sectors where economies of scale, branding, and advertising determine market share (e.g., petroleum products, chemicals, automobiles, food and beverages, and consumer goods). In mature sectors, while the technology underlying these industries may be diffused and codified, capital limitations and marketing capabilities can cause just a few MNEs to maintain a large share of the global market.

A majority of MNE activity is associated with firms from developed countries. As data on value adding and ownership are not readily available in a comparable form for all MNEs, most analyses rely on outward FDI data as a proxy for MNE activity. Approximately 87 percent of all outward FDI stock in 2004 was associated with MNEs whose ultimate beneficiary owner was based in the developed countries. Of the U.S. $1.4 trillion of outward FDI from developing countries, a vast majority some 90 percent is associated with 20 emerging market and transition economies and about 100 MNEs. These trends reflect the strong positive correlation between the level of economic development and the growth of firms from these economies with the technological and organizational capabilities to compete globally with the more established MNEs in world markets (UNCTAD 2006).

History of the MNE The MNE as a cross-border-coordinating, organizational entity (as it is known today) is a relatively new phenomenon that arose toward the end of the 19th century. During the 18th century, the large trading houses (state owned or state sanctioned) engaged primarily as traders, moving natural resources from the developing world to Europe and manufactures in the other direction. Although there were considerable foreign investments abroad, these were mainly associated with long-term capital flows the financing of debts, bonds, and other financial instruments required to support large-scale private and public projects in the

colonies. Firms were *international*, but not *multinational*. International business and economic activity was *extensive* in the sense that the value of goods and capital exchanged was considerable and involved numerous countries and actors. But it was not *intensive*, in that activities were largely not integrated across borders, with the possible exception of the large trading companies and other state-sanctioned de facto monopolies (Held et al. 1999). In addition, where private investments by individuals occurred, these were freestanding companies companies established by investors of a foreign nationality who were located in the host country, but not part of any firm located in the investor's home country. They were not strictly domestic firms, since their capital (and entrepreneurial basis) came from abroad, but they did not represent a direct investment since the foreign-located firm was not a subsidiary of another firm in another country. Jones (1996) notes that such investments were often associated with colonial interests. Indeed, even by 1914, the number of freestanding companies far outnumbered true MNE subsidiaries by a large order of magnitude. Both types of MNEs were predominantly in services and natural resource extraction. Most activities were *trade-supportive* in nature and involved little interdependence.

The first MNEs did not blossom until well into the 20th century, as the concept of large firms did not make its debut until late in the 19th century. The importance of size arose for several reasons, some technological. First, the large firm emerged about the time of the second industrial revolution, which featured innovations associated with mass-production techniques. These innovations particularly benefited more capital-intensive industries that were characterized by economies of scale and scope, such as refineries (both for sugar and petroleum), animal and vegetable oils, chemicals, iron, copper and aluminum (both primary processing and products), and packaging and processing of food products; by standardized machinery; and by the development of large-scale enterprises. Second, developments in transportation and communication made efficiently coordinating far-flung activities much more feasible.

Third, the brief period of free trade that had flourished in the 19th century was replaced after about 1880 by high trade barriers. FDI continued to be dominated by MNEs from the major European colonial powers Netherlands, the United Kingdom, and France and the United States, which was by this time firmly established as the technological and economic hegemon. These four countries accounted for about 87 percent of FDI stock in 1938 (Jones 1996).

Motives for MNE Investment There are four main motives for investment: to seek natural resources, to seek new markets, to restructure existing foreign production through rationalization, and to seek strategically related created assets. The first three represent motives that are primarily *asset-exploiting* in nature: the investing company's primary purpose is to generate economic rent through the use of its existing firm-specific assets. The last is a case of *asset-augmenting activity*, whereby the firm wishes to acquire additional assets that protect or augment its existing created assets in some way.

Typology of MNE Subsidiaries The literature on subsidiary development has greatly expanded over the last 20 years, evaluating the dynamics behind the evolution of subsidiary roles, beginning with the seminal work of Bartlett and Ghoshal (e.g., 1989). The strategic management literature on the role, dispersion, and development of subsidiaries is now well developed (see e.g., Pearce 1989; Birkinshaw and Hood 1998).

The nature of the activities undertaken by a subsidiary and its potential level of embeddedness in the host economy vary according to the level of competence of the subsidiary and the scope of its activities. The typology of subsidiaries can be analyzed according to these two scales. A typical value chain can be viewed from a "level of competence" perspective consisting of "strategic" and "operational" elements. Activities such as sales and manufacturing are operational in nature, while R&D centers and headquarters functions are strategic in nature. In general, strategic elements tend to be located close to locations regarded as important to the MNE. There is a close link between the influence of the subsidiary and the

strategic importance of its local environment. Strategic elements perform a critical role in a network of units, adding value through contributing their own expertise as well as by coordinating the flow of knowledge within the network. Second, there is considerable variation between subsidiaries in the scope of their activities, with certain subsidiaries performing single and specialized activities, and others performing a larger variety and of greater value.

Truncated miniature replicas. As their name implies, truncated miniature replicas are essentially a duplication of the parent firm, although not with the same scale of production and not all components of value adding activity. Typically, they do not undertake basic research but may modify and adapt products originally developed by the parent. Although truncated miniature replicas vary in the extent to which they are truncated, they tend to have a low or medium level of competence. Truncated miniature replicas tend to have an extensive market scope, in the sense that they have a large product range, but supply a limited and isolated market. Truncated miniature replicas tend to have a considerable autonomy in their activities, although the parent company exerts overall strategic control. This means, for instance, that the parent decides new additions to the product range. They are nationally responsive and, apart from a few advantages derived from being part of a MNE network such as lower cost of capital and technology, they are similar to other indigenous firms. Their primary motive is market-seeking and most often associated with import-substituting programs. The parent-affiliate relationship is weakly developed.

Rationalized affiliates. Rationalized affiliates are much more closely integrated into the MNE network. Their operations are motivated by efficiency seeking, aimed at optimizing costs over multiple locations. There is a strategic interdependence between the MNE network and the affiliate. Two types of rationalized affiliates can be distinguished: the rationalized production subsidiary and the world product mandate subsidiary. World product mandate subsidiaries maintain global or regional control over a particular product line or functional area and are designated centers of excellence. That is, strategic activities such as R&D and headquarters functions are included in the affiliate's responsibilities and it exerts control over other affiliates in the same region or worldwide. A rationalized production subsidiary is part of the MNE's global strategy and is engaged in the production of a particular value-adding aspect based on specific competitive advantages of the subsidiary relative to other subsidiaries. Its products are often intermediate goods, or products, or services complementary to other rationalized production subsidiaries.

Single-activity affiliates. Single-activity affiliates are a cross between truncated miniature replica and rationalized production subsidiary. On the one hand, they represent an extreme version of a truncated miniature replica, in that they undertake a single aspect of value-adding activity. On the other hand, such affiliates may in fact be part of a company's rationalized strategy: the comparative advantage of the location is best suited for such activities. These affiliates are not involved in decision making or strategic planning. Such affiliates typically tend to be engaged at the extremes of the value-adding chain. The first type is *trading affiliates*, engaged in trading activities, and, in the limit, in marketing and after-sales service. The second subcategory is *resource-extractive affiliates*, at the other end of the value chain, engaged in acquiring (primarily through extractive activity) scarce or otherwise valuable crude resources, and exporting these raw materials for use in other locations, whether by another affiliate or by an unrelated firm.

There is considerable variation between industrial sectors, individual MNEs, as well as host and home country factors. For instance, in the food and beverages sectors, subsidiaries are organized primarily as truncated miniature replicas. MNEs with greater international exposure and dependence on foreign markets are more inclined toward rationalized production subsidiaries or world product mandates. Subsidiary roles evolve over time, due both to internal, MNE-specific factors and to changing non-firm exogenous developments, including liberalization of markets and regional integration.

The changing external environment inevitably induces some changes in subsidiary roles. Once an MNE rationalizes the number of subsidiaries or re-organizes the activities across borders, the remaining and/or new units will likely experience changes in scope and areas of responsibility. Increases in scope can typically be found when the number of sub-sidiaries is rationalized or local conditions encourage localization of activities (Birkinshaw 1996). Simi-larly, the scope may be narrowed to focus on specific activities and build expertise within the selected area. Hence changes in scope are often related to both organizational and spatial considerations.

MNEs and Economic Development One of the key features of policy liberalization since the 1980s has been the desire to attract FDI as a means to ac-quire or improve technological capabilities through MNE activity. The role of the MNE as an additional source of capital and technology is one of the key features of this openness. The failure of protected industries in developing and developed countries to become competitive in global markets has high-lighted the limitations of the arm's-length technol-ogy transfer approach. Hence, in recent years, both governments and supranational organizations have increasingly come to focus on the role MNEs and FDI can play in development. This has been ac-companied by a lifting of many types of regulations that previously limited the role of FDI and MNEs.

MNEs have played and are likely to continue to play an important role in the structural upgrading of countries. However, the extent and pattern of these benefits is strongly dependent on the form of eco-nomic and social development desired by the host countries and on the policies of host governments in pursuing these goals. Although not the only means available, MNE spillovers are regarded as one of the most practical and efficient means by which indus-trial development and upgrading can be promoted (Narula and Dunning 2000).

While the *potential* for MNE-related spillovers is clear, as are the opportunities for industrial upgrad-ing therefrom, it is increasingly acknowledged that the nature, level, and extent of the benefits vary considerably. Even where MNEs do seek to transfer knowledge, they prefer to use technologies that are suited (first and foremost) to their own needs and the purposes for which they have made the investment. MNEs tailor their investment decisions to the exist-ing market needs and the relative quality of location advantages, especially skills and capabilities that the domestic economy has a comparative advantage in (Lall and Narula 2004).

Once the decision to enter a given market through FDI is taken, the kinds of activity and the level of competence of the subsidiary are also codetermined by the nature of the location advantages of the host location. While MNE internal factors such as their internationalization strategy, the role of the new lo-cation in their global portfolio of subsidiaries, and the motivation of their investment are pivotal in the structure of their investment, they are dependent on the location-specific resources available for that purpose.

The relative importance of the main motives of MNE investment partly reflects the stage of eco-nomic development (Dunning and Narula 1996). Least-developed countries tend to have mainly resource-seeking MNEs and countries at the catch-ing-up stage mostly market-seeking MNEs. Effi-ciency-seeking investments, with the most stringent capability needs, will tend to focus on the more in-dustrialized developing economies (though three or four decades ago they went to countries with rela-tively low capabilities, e.g., the electronics industry in Southeast Asia in the 1970s).

Not all affiliates offer the same spillovers to host economies. A sales office, for instance, may have a high turnover and employ many people, but its technological spillovers will be limited relative to a manufacturing facility. Likewise, resource-seeking activities like mining tend to be capital intensive and provide fewer spillovers than market-seeking man-ufacturing FDI. During import substitution, most MNEs set up truncated miniature replicas of their facilities at home. The extent of truncation, however, varied by host country. The most important deter-minants of truncation   and thus the scope of ac-tivities and competence of the subsidiary   were market size and local industrial capabilities (Dun-

ning and Narula 2004). Countries with small markets and weak local industries had the most truncated subsidiaries, often only single-activity subsidiaries (sales and marketing or natural resource extraction). Larger countries with domestic technological capacity (such as Brazil and India) had the least truncated subsidiaries, often with R&D departments.

With liberalization, MNE strategies on affiliate competence and scope have changed in four ways (Dunning and Narula 2004). First, there has been investment in *new affiliates.* Second, there has been *sequential investment* in upgrading existing subsidiaries. Third, there has been some *downgrading of subsidiaries*, whereby MNEs have divested in response to location advantages elsewhere or reduced the level of competence and scope of subsidiaries. Fourth, there has been some *redistribution of ownership* as the result of privatization or acquisitions of local private firms. In many, but certainly not all, cases this also led to a downgrading of activities.

MNEs are taking advantage of liberalization to concentrate production capacity in a few locations, exploiting scale and agglomeration economies, favorable location, and strong capabilities. Some miniature replicas have been downgraded to sales and marketing affiliates, with fewer opportunities for spillovers. Countries that receive FDI with the highest potential for capability development are those with strong domestic absorptive capacities.

MNEs transfer technology to local firms in four ways: backward linkages, labor turnover, horizontal linkages, and international technology spillovers. Studies of backward linkages have identified various determinants, including those internal to MNEs and those associated with host economies. The ability of the host economy to benefit from MNE linkages has been found to depend crucially on the relative technological capabilities of recipient and transmitter: the greater the distance between them, the lower the intensity of linkages.

Again, MNE motives and strategies matter. Market-oriented affiliates generally purchase more locally than export-oriented firms because of lower quality requirements and technical specifications.

MNEs create more linkages when they use intermediate goods intensively, communication costs between parent and affiliate are high, and the home and host markets are relatively similar in terms of intermediate goods. Affiliates established by mergers and acquisitions are likely to have stronger links with domestic suppliers than those established by greenfield investment — which is a subsidiary that has been established from scratch, as opposed to the acquisition of an existing facility — since the former may find established linkages that they are likely to retain if the linkages are efficient. Linkages vary significantly by industry. In the primary sector, the scope for vertical linkages is often limited, due to the use of continuous production processes and the capital intensity of operations. In manufacturing, the potential for vertical linkages is broader, depending on the extent of intermediate inputs to total production and the type of production processes.

Furthermore, the individual MNE's choice of mode of entry — whether to create a wholly owned subsidiary or engage in a joint venture or minority ownership — plays a significant role in the extent that spillovers and externalities accrue to host locations and firms. For instance, MNEs may be more likely to transfer sophisticated technologies and management techniques to wholly owned subsidiaries than to partially owned affiliates.

MNEs and unrestrained flows of inward FDI may well lead to an increase in productivity and exports, but they do not necessarily result in increased competitiveness of the domestic sector or increased industrial capacity, which ultimately determines economic growth in the long run. FDI *per se* does not provide growth opportunities unless a domestic industrial sector exists that has the necessary technological capacity to profit from the externalities from MNE activity. To put it simply, FDI is not a sine qua non for development (Lall and Narula 2004).

Political Economy, International Regulation, and MNE The role of foreign capital and foreign capitalists — as MNE activity was referred to in the literature prior to the 1950s — has historically been seen with some suspicion, due in part to the association of MNEs with imperialism, consequential

dependent development, and political and economic intervention by the home countries of MNEs on their behalf. MNE activity has often been inhibited as a result of protectionism, techno-nationalism, and the support of domestic industry against the possible negative effects of MNEs on domestically owned firms.

Economic liberalization whether voluntary, imposed as a condition for lending by international institutions, or as a precondition for membership of a regional integration program has led to a sea change in the policies of many countries. However, although policies may have changed, the underlying attitudes have not. The politics of state-MNE relationships have remained strongly associated with national interests, domestic interest groups, and the importance of protecting the state as a political and economic sovereignty. Supranational agreements such as the North American Free Trade Agreement and the agreements that formed the European Union and the World Trade Organization have reinforced, accelerated, and created standardized regulation for economic activity, acting as a virtuous circle with regard to economic integration that had been occurring as a matter of course.

See also foreign direct investment (FDI); globalization

FURTHER READING

Bartlett, Christopher, and Sumantra Ghoshal. 1989. *Managing Across Borders: The Transnational Solution.* Cambridge, MA: Harvard Business School Press. A strategic management approach to how MNEs are best organized and run.

Birkinshaw, Julian. 1996. "How Multinational Subsidiary Mandates Are Gained and Lost." *Journal of International Business Studies* 27 (3): 467 95. Explains how the different subsidiaries can have different roles.

Birkinshaw, Julian, and Neil Hood. 1998. "Multinational Subsidiary Evolution: Capability and Charter Change in Foreign Owned Subsidiary Companies." *Academy of Management Review* 23 (4): 773 95. Paper that examines how subsidiary mandates evolve over time.

Caves, Richard. 2007. *Multinational Enterprises and Economic Analysis.* Cambridge: Cambridge University Press. An authoritative text that takes a primarily economics view.

Dunning, John. 1993. *Multinational Enterprises and the Global Economy.* Wokingham, UK: Addison Wesley. A classic reference that covers the subject from economics, management, and international business perspectives.

Dunning, John, and Rajneesh Narula. 1996. *Foreign Direct Investment and Governments.* London: Routledge. A study of how MNE activity, economic structure, and economic development are interrelated.

. 2004. *Multinationals and Industrial Competitiveness: A New Agenda.* Cheltenham, UK: Edward Elgar. Collection of essays that examines current issues in international business.

Held, David, Anthony McGrew, David Goldblatt, and Jonathan Perraton. 1999. *Global Transformations: Politics, Economics, and Culture.* Cambridge: Polity Press. An authoritative study of globalization, linking the political, sociological, and economics perspectives.

Jones, Geoffrey. 1996. *The Evolution of International Business.* Routledge: London. A text that takes a business history approach to MNEs.

Lall, Sanjaya, and Rajneesh Narula. 2004. "FDI and its Role in Economic Development: Do We Need a New Agenda?" *European Journal of Development Research* 16 (3): 447 64. Covers the main issues of FDI and development in a post WTO world.

Narula, Rajneesh. 2003. *Globalization and Technology.* Cambridge: Polity Press. Analysis of the interdependencies between economic globalization, innovation, technology, and industrial policy.

Narula, Rajneesh, and John Dunning. 2000. "Industrial Development, Globalization, and Multinational Enterprises: New Realities for Developing Countries." *Oxford Development Studies* 28 (2): 141 67. Looks at the effect of MNEs in industrial development.

Pearce, Robert. 1989. *The Internationalization of Research and Development by Multinational Enterprises.* Macmillan: London. How MNEs have internationalized their R&D activities.

UNCTAD. 2006. *World Investment Report 2006.* Geneva: United Nations. Part of an annual series of reports by UNCTAD with relevant statistical and anecdotal data on MNE activity.

Wilkins, Mira. 2001. "The History of the Multinational Enterprise." In *The Oxford Handbook of International Business*, edited by Alan Rugman and Thomas Brewer. Oxford: Oxford University Press, 3 35. A historical overview of the evolution of the MNE.

RAJNEESH NARULA

■ multiple currencies

When multiple currencies circulate in a country or region, it is possible to pay for most goods and services there with more than one currency. This entry focuses on government-issued currencies. The entry on *currency competition* discusses the interaction of government-issued money with privately issued monies.

Historically, it has been common for multiple currencies to circulate in a given area. In medieval Europe, for example, and in the United States before the Civil War, coins from many countries circulated side by side. More recently, multiple currencies have circulated in developing countries, notably in South America and Asia, and in some areas of developed countries. In certain parts of England, for instance, most goods and services can be purchased with euros and pounds sterling.

In developed countries, multiple currencies circulate mainly for convenience, to attract or accommodate the business of merchants or tourists. In developing countries in which multiple currencies circulate today, the purpose may be different. The possibility of using dollars or euros protects people in these countries from the cost of inflation of their domestic currencies and provides a way to discipline their monetary authorities.

Benefits of Multiple Currencies Central banks today issue fiat money, or money that is intrinsically worthless and is not backed by an asset or a commodity at the central bank. Fiat money has value because people anticipate that others will agree to exchange goods and services for the money. It is generally well understood that issuance of fiat money suffers from a time inconsistency problem, in which the government chooses to change its policy from that promised earlier, often after other actors have made decisions or investments based on the previously announced policy. The monetary authority, or the government that controls it, has an incentive to issue too much money because the cost of producing an extra unit of money is lower than the value of the money.

Businesses and consumers can choose to trade in another currency, one that they expect will retain its value. This gives them some protection from the inflation tax, a term that refers to the erosion of purchasing power. For this reason, one would expect multiple currencies to circulate in countries that have suffered from high inflation. Further, one would expect that the foreign currency that circulates is issued by a central bank with a reputation for maintaining the value of its currency.

Over the last 20 years, the Federal Reserve's record of keeping inflation low in the United States has been much better than that of a number of developing countries. Moreover, the institutional setting in which the Federal Reserve operates suggests that the U.S. dollar is unlikely to suffer from high inflation in the future. This is also true of other currencies, such as the euro. Hence, it is not surprising that the dollar or the euro often circulate in other countries.

The circulation of multiple currencies can also help discipline a monetary authority because it reduces the incentives to set high inflation. The easier it is for businesses and consumers to trade with another currency, the more constrained the monetary authority is. In some extreme cases, the local currency can disappear entirely and be replaced by a foreign currency. Because multiple currencies can serve to discipline a monetary authority, they can also be used as a commitment device in a country where institutions make it difficult for the monetary authority to commit to low inflation. This means that the impetus for multiple currencies may come from a government, rather than consumers and businesses. In such a case, the government aims to signal that it wants the monetary authority to refrain from implementing high inflation.

Costs of Multiple Currencies Although multiple currencies have benefits, they also have a number of costs. An important cost is the loss of flexibility in the conduct of monetary policy. It is generally believed that central banks can help smooth the fluctuations affecting an economy with the judicious use of monetary policy. In a recession, an expansionary monetary policy can stimulate the economy; in an expansion, a tight monetary policy can prevent the economy from growing too fast.

Multiple currencies reduce the central bank's ability to stimulate the economy with an expansionary policy that increases the rate of growth of the money supply. Moreover, monetary policy is less effective since it affects only a fraction of the economic activity. The central bank is unable to stabilize the local economy as well as it could if multiple currencies did not circulate. The consequence is volatility that hurts businesses and consumers.

Of course, the constraint imposed on monetary policy by multiple currencies is partly by design. The benefits of multiple currencies arise because the monetary authority has a tendency to choose a monetary policy that is too loose. There is a trade-off between reducing the incentives of the monetary authority to choose high inflation and giving it flexibility to stimulate the economy. Multiple currencies can be socially desirable if the costs of high inflation are higher than the benefits of stabilization policy. This is likely to be the case when the monetary authority finds it especially difficult to implement low inflation.

The use of multiple currencies also reduces flexibility in other areas, such as the ability of the central bank to implement lender-of-last-resort policies. In cases of financial crises, the central bank can sometime enhance the stability of the financial system by temporarily providing a large amount of liquidity. The effectiveness of this kind of policy is likely to decline in the presence of multiple currencies. This is especially true if the financial system relies heavily on the foreign currency. For example, a lender-of-last-resort policy to prevent bank runs is less effective if a large fraction of deposits at banks are denominated in dollars. An additional cost of multiple currencies is the inconvenience experienced by businesses and consumers when they have to deal with several currencies, such as keeping track of inventories for each currency.

Need to Weigh the Costs and Benefits Multiple currencies protect consumers from the inflation tax if their monetary authority is unable to keep inflation low. They impose discipline on the monetary authority and can be used as a commitment device to signal an intention to keep inflation low. Multiple currencies also reduce the ability of a country's monetary authority to smooth fluctuations, however, and can reduce the effectiveness of lender-of-last-resort policies. Because of these costs, multiple currencies are likely to circulate only in countries with a history of high and/or volatile inflation rates.

See also banking crisis; currency competition; currency substitution and dollarization; discipline; dollar standard; dominant currency; Federal Reserve Board; financial crisis; lender of last resort; reserve currency; seigniorage; time inconsistency problem; vehicle currency

FURTHER READING

Martin, Antoine. 2006. "Endogenous Multiple Currencies." *Journal of Money, Credit, and Banking* 38: 247 64. Formalizes many of the arguments described in this entry and provides a number of references.

ANTOINE MARTIN

■ Mundell-Fleming model

The Mundell-Fleming model, named for J. Marcus Fleming (1911 76) and Robert A. Mundell (born 1932), in its most familiar form is the "internationalized IS-LM model," a diagrammatic representation of the way in which a small, open economy responds to shocks that take the form of internal policy initiatives. The innovation in this approach is that the exchange rate regime is a crucial determinant of the policies' effects on economic variables, and specifically on the level of gross domestic product (GDP) generated by the country under analysis. In two particular cases, contrary to expectations, these initiatives do not have an impact on output in the

small, open economy in the perfect capital mobility setting. This is true both for monetary policy under fixed exchange rates and for fiscal policy under flexible exchange rates. For these situations the model shows that the initiatives are ineffective for influencing the level of output and employment.

The clear implication is that policymakers should be aware of the impotence of these initiatives and not rely on them in a fruitless attempt to stabilize the economy. Canada was faulted in the late 1950s for depending on expanded government expenditures to end a stubborn recession. Fleming's ineffectiveness result demonstrates that the impact of Canada's fiscal policy was vitiated because the Bank of Canada allowed the price of its dollar to be determined in the financial markets, rather than fixing it at a given value. The clumsy conduct of monetary policy by the Bank of Canada at the time was particularly unfortunate because this flexible exchange rate regime enhanced the potency of that policy. Although these observations were articulated by Fleming and Mundell a few years after these events, the intuitive feeling at the time was that something was amiss at the Bank of Canada, and dissatisfaction with its policies lay behind the impeachment of its governor in 1961.

The Hallmarks of the Mundell-Fleming Model
The Mundell-Fleming model is based on a small group of papers, including Fleming (1962) and Mundell (1963). Others, such as Mundell (1960, 1961, and 1962), at times have been incorporated into the rubric, but doing so risks jeopardizing the key insights derived from the central papers. Some of Mundell's papers (e.g., Mundell 1960) present models designed to generate comparative static results that are independent of the nature of the exchange-rate regime in effect. Others (e.g., Mundell 1961) find comparative static results that contradict the ineffectiveness conclusions of the central papers. A key reason for this is that many of these papers assume that monetary policy can be gauged in terms of the level of interest rates rather than the quantity of money outstanding. Still others (e.g., Mundell 1962) are phrased entirely in dynamic terms rather than in the comparative statics one finds in the central papers.

Readers of these central papers face daunting challenges, as the mathematics presented is either elliptical (Mundell) or complicated (Fleming). But what is distinctive about these papers is that they employ a definition of monetary policy that maintains its coherence when capital mobility becomes perfect, and that they deal consistently with the flexible exchange rate case.

Diagrammatic Representation The Mundell-Fleming model is portrayed in figure 1. On the axes are two variables central to macroeconomic performance: on the horizontal axis, Y measures domestic output (GDP) produced in the small, open economy; on the vertical axis, r is the level of domestic interest rates. Two of the loci are familiar from Hicks's (1937) classic article: XX, representing equilibrium in the domestic goods market; and LL, representing equilibrium in the domestic assets market.

The XX locus is negatively sloped because some expenditures are sensitive to interest rates. With a higher interest rate, such expenditures are reduced, and other things being equal, this causes output to decrease. One of the variables held constant in drawing this curve is the value of the exchange rate. Defining the exchange rate as the domestic-currency price of one unit of foreign currency, we note that a rise in the exchange rate's value tends to raise the price
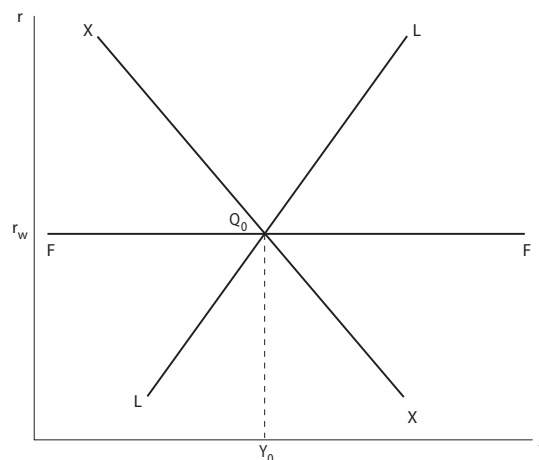
Figure 1
The Mundell Fleming model

of foreign goods relative to domestic goods. The resulting lower relative price for domestic goods increases demand for them and, according to the Keynesian view of supply, output responds to this rise in demand at constant prices. With such a change, the XX locus would shift to the right. We proceed with the assumptions that the quantity of money in this small economy does not have a direct impact on the goods market equilibrium condition, and that the domestic-currency price level has a value which is given.

The LL locus is positively sloped, being drawn for a given quantity of money and given values of domestic-currency prices. Everywhere along this locus the quantity of money demanded is constant, consistent with its given supply, so that for these points there is equilibrium in the domestic financial assets market. The reason is that an increase in interest rates causes the quantity of money demanded to be lower, since now bonds, an alternative form in which to hold funds, provide a higher yield. If this is matched by an increase in output, then the resulting higher volume of transactions necessitates an increase in the quantity of money held. In the right proportions, these two influences on money demand exactly offset each other and demand is back at its original (equilibrium) value. An increase in the money supply will cause a rightward shift in the LL curve.

The argument concerning the LL locus, for the simple case in which domestic-currency prices are taken to be constant, can be represented algebraically as follows:

$$M = l(r) \cdot Y \qquad (1)$$

This specification is consistent with the one in Fleming, in that it assumes a unitary income elasticity of demand for money, where l is the inverse of velocity. If l depends negatively on the interest rate, then a rise in r must be matched by a rise in Y in order to keep the product of the two factors on the right-hand side of this equation constant. This is necessary in order to maintain equality to the given value of the money supply, M, in this equation. This specification employs the assumption of the standard model that demand for money is independent of the exchange rate.

As noted, these two curves are identical to those in Hicks (1937). Hicks's model is the basis for the simplified presentations of conventional macroeconomic theorizing, so little further discussion is needed here, except to note that the XX locus is relevant for equilibrium in the domestic goods market alone. No separate analysis needs to be carried out for the market for imports, since this small economy is able to obtain all the foreign-produced goods it demands.

The final curve, FF, was added by Mundell to represent equilibrium in the foreign exchange market. In the perfect capital mobility case shown, this locus is horizontal: if domestic bonds are perfect substitutes for foreign bonds, then their rates of return must be equated. No matter what the level of domestic output, the yield on domestic bonds must equal the yield available from holding foreign financial instruments, denoted by r_w.

The economy is assumed to start in equilibrium at point Q_0, where there is clearing (so that supply and demand are equated) of all the markets analyzed here.

Policy Shocks This diagram can be used to derive the effects of monetary policy (changes in the quantity of money, causing the LL locus in figure 1 to shift) and fiscal policy (changes in the level of expenditures by government, which shift the goods-market clearing condition, XX in figure 1, independent of displacements caused by exchange rate changes) under fixed and flexible exchange rates. These are the central applications of the Mundell-Fleming model. The difference between these two exchange rate regimes is that under fixed exchange rates the value of the exchange rate is given, and the central bank, in pegging the exchange rate at a particular value, uses changes in the quantity of money to maintain this value. That is, for this regime the money supply becomes a variable that responds to shocks, while the value of the exchange rate does not change. In contrast, for flexible exchange rates the quantity of domestic money outstanding is treated as a given, and the value of the exchange rate moves reactively in order to reequilibrate the system.

The effects of expansionary monetary and fiscal policy actions can be stated easily when the exchange

rate is flexible: in that case fiscal policy has no impact on output, whereas monetary policy has the effect to which we are accustomed. These conclusions derive directly from equation (1).

If the money supply is held constant in the face of an increase in government expenditures, then the level of output does not change. With such a shock, M is fixed, and the value of l, as well, is constant, because the perfect capital mobility assumption views the interest rate as unchanging. If both M and l are given, then Y cannot change, as equation (1) shows. This is the famous "crowding out" or "ineffectiveness" result Fleming derived, which Mundell repeats.

In contrast, changes in the money supply, due to monetary policy initiatives, cause output to change in proportion to those changes. This conclusion is an unusual but straightforward implication of the Fleming specification of money demand. Once again the derivation depends on the assumption of perfect capital mobility, which causes the value of l to be given because it is a function of only the interest rate, whose value is set by world market conditions.

These results are portrayed in figure 2. An increase in government expenditure shifts the XX locus to the right by a multiple of the amount of increase. With its original position shown as XX, this shift generates a new position for this locus shown by X'X', drawn as a
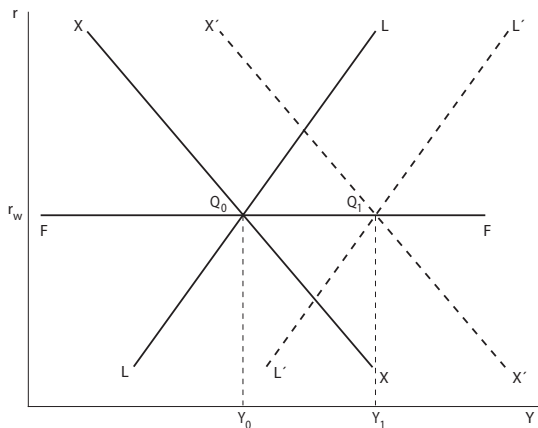
dotted line to emphasize that this is the new displaced position of that curve.

If the exchange rate is fixed by the central bank, then it must provide the larger quantity of money that is required at an unchanged interest rate. As this happens, the LL locus shifts to the right; this process continues until the locus attains the position shown as L'L'. The new equilibrium is at point Q_1 where output and the money supply have increased (with output now having a value shown by Y_1 whereas its previous value was Y_0), and the values of both the interest rate and the exchange rate are unchanged.

In contrast, the same increase in government expenditures will have no effect on output and employment if the central bank is pursuing a flexible exchange rate regime. For this case, the money supply is held constant, and the value of the exchange rate must change in order to clear the various markets. Since money demand does not depend on the value of the exchange rate, its changing value does not alter the position of LL, which continues to run through point Q_0. Instead, it is the XX locus that shifts back to its original position, as the value of the exchange rate falls, in order to reequilibrate all of these markets. Since the equilibrium under this exchange rate regime does not move, it is clear that any increased government expenditures come at the expense of exports and imports. There is complete crowding out of such expenditures, so the composition of expenditures on domestic output changes, but its total quantity does not.

This idea   that with high capital mobility, fiscal policy has a more limited impact on output under flexible exchange rates than its does under fixed   is usually associated with Rhomberg (1964), who did the original empirical work on the Canadian experiment with flexible exchange rates in the 1950s. He appears to have inspired Fleming's interest in this topic, as the Canadian experience seemed to suggest that flexible exchange rates vitiated fiscal policy. Mundell, as well, credits Rhomberg's work in this area.

The argument is reversed for monetary policy. In this case the increase in the money supply is the given impulse that moves the money market equilibrium locus to position L'L'. Does this locus maintain this position, setting up an equilibrium level of output

Figure 2
The Mundell Fleming model: Changes in government expenditures and money supply

equal to Y_1, or does its revert to its original position, so that output remains at Y_0? The answer depends again on the exchange rate regime.

For fixed exchange rates, the quantity of money in the domestic economy is used to peg the value of the exchange rate. If that value is not changed, then there is no reason why the equilibrium quantity of money should change either. This means that the quantity of money quickly reverts to its former value (or perhaps does not move at all). The end result is that output is not affected by monetary policy in the fixed exchange rate case.

The quantity of money is a responding variable under a fixed exchange rate regime. So there is the logical problem that the quantity of money cannot legitimately be changed in an arbitrary fashion. As McCallum (1996) noted, one should describe this Hume impotency conclusion as demonstrating not the ineffectiveness of monetary policy, but rather its unavailability.

For flexible exchange rates, there is no commitment to keeping the value of the exchange rate at a particular peg. As a result, the decision to increase the quantity of money causes the money market locus to move to the dotted position shown, $L'L'$. In that case the value of the exchange rate must rise to clear the various markets. As domestic currency depreciates against foreign denominations, the goods market locus shifts over to the dotted position, $X'X'$. Output in this case increases to Y_1 showing that monetary policy is effective in influencing the level of domestic economy activity in the small country case.

Coining of the Expression The central articles in the Mundell-Fleming canon were written in the early 1960s. The expression *Mundell-Fleming model* was introduced, at the earliest, five years later. Soon after the term became part of the everyday vocabulary of researchers in international finance, but it was only after 1975 that the expression was used in print. Although many economists have used the alphabetical ordering of the names in the expression (Kenen 1985; Boughton 2003), and some have referred to it as just the "Fleming model" (Cooper 1976), *Mundell-Fleming model* came into common use by economists in the later 1970s and 1980s. During this period even significant extensions of the basic framework were still identified as being part of the Mundell-Fleming model (Marston 1985; Frenkel and Razin 1987).

The economists Maurice Obstfeld and Kenneth Rogoff (1996), in their monumental graduate textbook in international finance, refer to the Mundell-Fleming-Dornbusch model, and they do not use the conventional name. In the 1990s, Obstfeld and Rogoff developed the New Open Economy Macroeconomics (NOEM) model, which has the potential of supplanting the standard model as the basic benchmark formulation of the ideas and mechanisms of international financial analysis.

See also capital mobility; exchange rate regimes; impossible trinity; New Open Economy Macroeconomics; quantity theory of money; Swan diagram

FURTHER READING

Boughton, James. 2003. "On the Origins of the Fleming Mundell Model." *IMF Staff Papers* 50 (1) (April): 1 9. Analyzes the writing of Fleming's 1962 contribution to the Mundell Fleming model and concludes that Fleming's work was substantially independent of Mundell's.

Cooper, Richard N. 1976. "Monetary Theory and Policy in an Open Economy." *Scandinavian Journal of Economics* 78 (2) (June): 146 63. Provides an analysis of financial policies in the open economy and calls the conventional framework at the time "the Fleming model."

Fleming, J. Marcus. 1962. "Domestic Financial Policies under Fixed and under Floating Exchange Rates." *IMF Staff Papers* 9 (4) (November): 369 79. Considered to be the classic contribution by Fleming, this paper reports both verbally and mathematically the ineffectiveness result for fiscal policy, and the proportionality result for monetary policy for the flexible exchange rate, perfect capital mobility case.

Frenkel, Jacob A., and Assaf Razin. 1987. "The Mundell Fleming Model a Quarter Century Later." *IMF Staff Papers* 34 (4) (December): 567 620. Characterized by the authors as a survey piece, this paper presents a complicated two country model of the world economy in the perfect capital mobility setting. Since the analysis specifies the asset markets in a portfolio balance manner, the connection between this framework and

what is usually termed the Mundell Fleming model is unclear.

Hicks, John. 1937. "Mr. Keynes and the 'Classics': A Sug gested Interpretation." *Econometrica* 5 (2) (April): 147 59. The diagram included in this article has been the basis for most textbook presentations of closed economy macroeconomic theory right up to the present. The longevity of this framework has surprised both its pro ponents and its many critics.

Kenen, Peter B. 1985. "Macroeconomic Theory and Policy: How the Closed Economy Was Opened." In *Handbook of International Economics*, vol. 2, edited by Ronald W. Jones and Peter B. Kenen. North Holland, Amsterdam, 626 77, chapter 13. A survey piece that covers the tra ditional way of thinking about the development of the open economy macroeconomic model, using a number of distinct models and different modes of analysis.

Marston, Richard C. 1985. "Stabilization Policies in Open Economies." In *Handbook of International Economics*, vol. 2, edited by Ronald W. Jones and Peter B. Kenen. North Holland, Amsterdam, 859 916, chapter 17. Presents a portfolio balance analysis of the effects of standard macroeconomic shocks. The author uses a complex model and the stock specification of the asset markets. He describes his framework as a "modified Mundell Fleming model."

McCallum, Bennett T. 1996. *International Monetary Eco nomics*. New York: Oxford University Press. A standard presentation of open economy macroeconomic topics. The author concludes that Mundell's 1962 and 1963 papers are his central contributions to the Mundell Fleming model.

Mundell, Robert A. 1960. "The Monetary Dynamics of International Adjustment under Fixed and Flexible Ex change Rates." *Quarterly Journal of Economics* 74 (2) (May): 227 57. Includes a model specifically designed so that the comparative static responses of the values of real variables are independent of the nature of the ex change rate regime. Despite its title, this article includes neither the money supply nor the exchange rate in the central mathematical model, which one finds in its ex tensive appendix.

———. 1961. "Flexible Exchange Rates and Employment Policy." *Canadian Journal of Economics and Political Science* 27 (4) (November): 509 17. Although the fixed exchange rate case is not treated explicitly in this pre sentation, its conclusions appear to conflict with those published in the famous paper two years later. In par ticular, monetary policy seems to have potency under fixed exchange rates, and fiscal policy is reported as being effective under flexible exchange rates.

———. 1962. "The Appropriate Use of Monetary and Fiscal Policy for Internal and External Stability." *IMF Staff Papers* 9 (1) (March): 70 77. Presents in a dynamic form an argument found at the end of another Mundell article, which deals with the appropriate response to a situation of internal and external imbalance. This article has been the basis for the author's thinking about the "policy mix" and supply side economics.

———. 1963. "Capital Mobility and Stabilization Policy under Fixed and Flexible Exchange Rates." *Canadian Journal of Economics and Political Science* 29 (4) (No vember): 475 85. Mundell's most frequently cited pa per. Reports results like those in Fleming (1962), espe cially for the flexible exchange rate, perfect capital mobility case. Juxtaposes the Fleming and Hume inef fectiveness results and presents the famous "inter nationalized IS LM" diagram.

Obstfeld, Maurice, and Kenneth Rogoff. 1996. *Foundations of International Macroeconomics*. Cambridge, MA: MIT Press. A monumental textbook presenting the state of the subject at the time of publication. The analysis focuses on current account dynamics for its real coverage, and on the flexible exchange rate case in its monetary topics. It has an extensive treatment of the Dornbusch "over shooting" framework, which the authors call the "Mundell Fleming Dornbusch model"

Rhomberg, Rudolph R. 1964. "A Model of the Canadian Economy under Fixed and Fluctuating Exchange Rates." *Journal of Political Economy* 72 (1) (February): 1 31. Finds that government expenditures are likely to lead to an appreciation of domestic currency in the in termediate run. This argument may have inspired Fleming to write down his model so as to explain how the effects of fiscal policy were vitiated in Canada in the late 1950s.

RUSSELL S. BOYER

N

■ **national treatment**

See nondiscrimination

■ **neoliberalism**

See dependency theory

■ **New Economic Geography**

Abundance of natural resources, proximity to natural means of communication, and climatic conditions vary from place to place. These variations, however, provide only a partial explanation of the pronounced differences in development that exist even between areas that have fairly similar characteristics. Other forces inherent to the functioning of economic interactions are able to cause uneven development even across otherwise identical places.

While geographers have proposed a rich list of such forces, since the mid-1990s the work of regional scientists and urban economists called "New Economic Geography" (NEG) has gained momentum in mainstream economics. What distinguishes NEG from other approaches is its focus on the localized market rather than nonmarket interactions within a "general equilibrium" framework that emphasizes the determination of good and factor (capital, labor and land) prices as well as the importance of economywide budget constraints.

Location Decision The cornerstone of NEG is the location decision of the firm. This is a nontrivial practical problem provided that the shipment of goods and factors across space as well as the fragmentation of production affect the firm's total costs due to the presence of trade obstacles and plant-level scale economies, respectively. It is also a well-defined theoretical problem provided that the firm has some market power, because plant-level scale economies are incompatible with a perfectly competitive equilibrium. Increasing returns to scale at the plant level and costly transportation generate an economic trade-off between the proximity to dispersed customers and suppliers on the one hand, and the concentration of production in a few large plants on the other. Proximity reduces transportation costs but requires the firm to fragment production across many small plants. Concentration allows the firm to exploit scale economies but increases its transportation costs. Accordingly, the firm will prefer to run several small dispersed plants when returns to scale are weak and transportation costs are large. It will prefer to concentrate production in a few large plants when returns to scale are strong and transportation costs are small. Other things being equal, transportation costs and scale economies also imply that the firm will be attracted to markets with large local demand. This is sometimes called the home market effect, whereby firms tend to solve the trade-off between proximity and concentration by locating in the larger market and serving the smaller one from there rather than vice versa. The more significant the home market effect, the lower the transportation costs and the stronger the scale economies.

The location decision of the firm is made more complex by the presence of competitors because its

geographical position with respect to them affects the market power that necessarily stems from plant-level economies of scale. Generally speaking, firms have market power when they do not take market prices as given, as perfectly competitive firms would, and therefore trade quantity against price in making their profit-maximizing decisions. This price-making behavior is the essence of imperfectly competitive market structures that arise when competitors are a small group (oligopoly) or when they are a large group offering differentiated products (monopolistic competition). Although NEG models focus mainly on the latter, in both cases location turns out to be a crucial decision variable for profit maximization because it allows firms to increase market power and thus profits by carefully choosing their production sites.

In terms of social welfare, the presence of market power implies that prices, on which households and firms base their consumption, production, and location decisions, do not fully reflect the corresponding social values. Thus market interactions generate "side effects" for which no quid pro quo is paid. Side effects that are associated with market transactions are called "pecuniary externalities" and imply that the location decision of the firm does not fully take into account its effects on customers, suppliers, and competitors. Approaches other than NEG prefer to focus instead on "technological externalities." These are independent of any market interaction; they materialize through sheer physical proximity. Because they are the outcome of non-market interactions, by definition no quid pro quo is paid for them. As an example, consider an industry ("upstream") that supplies intermediates to another industry ("downstream"). A positive pecuniary externality for the downstream industry is the fall in intermediate input prices due to the increase in upstream competition triggered by the entry of a new technologically advanced supplier; a positive technological externality is the increase in productivity that other upstream suppliers may experience through informal knowledge transmission (spillover) generated by their proximity to their new technologically advanced rival. Although both concepts

have their own merits, the logical advantage of pecuniary externalities lies in the possibility of relating their emergence to a set of well-defined microeconomic parameters. So far this has proven to be quite difficult in models based on technological externalities because these remain mostly black boxes. Models with pecuniary or technological externalities do, however, share similar welfare implications: the free-market economic landscape is inherently inefficient and appropriate public intervention is generally needed.

Cumulative Causation No matter whether through market or nonmarket interactions, the attractiveness to a firm of alternative production sites depends on where other firms locate. This may activate a mechanism of cumulative causation among firms' location decisions through which firms' interactions may alter the economic landscape implied by natural resources, natural means of communication, and climatic conditions.

To understand how cumulative causation might work, let us go back to the previous example of upstream and downstream industries. For simplicity, consider a stylized economy made of two initially identical locations, A and B. Both host similar production chains in which there are three vertically linked activities: intermediate production, final production, and consumption. For simplicity, assume that final production uses only intermediate inputs, intermediate production employs only labor, and workers are the only source of final demand and are geographically immobile. If for any reason a new firm starts producing intermediates in A, it will increase local labor demand and intermediate supply. Due to excess demand and supply, respectively, local wages will go up while local intermediate prices will fall. This is bad news for the other local intermediate producers (market crowding effect due to competitor proximity). It is good news, however, for local final suppliers since they experience falling production costs and higher demand by richer workers. As new final producers are induced to enter the market in A, the expansion of final production will feed back into stronger intermediate demand so that intermediate suppliers will benefit (market expansion effect due to

customer proximity). Clearly, when the latter effect dominates the former, both final and intermediate firms will end up being agglomerated in A. Through these processes, cumulative causation among firms' location decisions can generate persistent differences even among places that are initially identical.

Both labor mobility and capital accumulation tend to reinforce the market expansion effect through the additional income they generate. Intermediate entry expands labor demand and final production, which itself feeds back into stronger demand and hence increased production of intermediates. If labor is geographically mobile, the associated rise in wages in location A attracts workers from B. As these migrate, demand in the final market increases, generating a second cycle of cumulative causation between firms' and workers' location decisions. Analogously, the expansion of intermediate and final markets increases firm profitability, thus fostering investment in capital accumulation and innovation. The additional income generates a third cycle of cumulative causation.

As a parallel and opposite cycle of income and expenditures contraction arises in location B, even small and transitory location-specific shocks can give rise to large, permanent geographical imbalances. Thus there is a priori great flexibility on where particular activities locate. Once the agglomeration process has started, however, spatial differences take shape and become quite rigid, a process known as putty clay geography.

Trade Liberalization Although cumulative causation is present in many models of regional development with localized externalities, the crucial contribution of NEG is that its microeconomic foundations allow the evolution of the spatial landscape to be related to measurable microeconomic parameters. A first example is the prediction that agglomeration is more likely to take place in sectors where increasing returns are more intense and product differentiation is more pronounced. The reason is that in those sectors the market crowding effect is weaker as more intense returns to scale reduce the number of active competitors and more product differentiation reinforces market power. A second

example highlights what is arguably the most celebrated insight of NEG models: the impact of freer trade on the economic landscape.

NEG argues that the level of trade barriers affects the balance among market expansion and market crowding. Consider a situation in which some factors such as land are immobile and ubiquitous. Assume also that the incomes they generate are spent locally. With extremely high trade barriers, firms will be dispersed as scattered customers can be reached only through local production. As trade barriers fall, however, it becomes economically viable to serve scattered customers remotely. This unleashes cumulative causation so that regional imbalances may arise endogenously through the location decisions of firms and mobile factors. As trade barriers keep on falling, however, firms' geographical positions gradually become immaterial in terms of product market interactions. Eventually what is left is only the dispersion force due to the higher prices of immobile factors in the crowded place. As agglomeration unfolds in the process of moving toward low-level trade barriers, trade liberalization has a nonlinear effect on the spatial concentration of economic activities by promoting agglomeration at early stages and dispersion afterward.

Multilocation Economies The results surveyed so far are typically derived in stylized economies in which there are only two locations. A more realistic description of the geographical space entails investigating the behavior of multilocation economies. Considering more than two locations does not affect the agglomeration effect of more intense scale economies and stronger market power. It also does not affect the nonlinear impact of trade barriers on the spatial concentration of economic activities as agglomeration is still likelier for intermediate obstacles to trade. Nonetheless, with more than two locations, relative market sizes are not enough anymore to determine where firms cluster. The relative positions of locations within the network of trading relations are also crucial: locations that have better accessibility will be more likely to attract firms.

In two-location models the home market effect implies that an expansion of expenditures in a certain

location causes a more than proportionate increase in local supply thanks to the influx of new firms. This effect, which is at the source of cumulative causation, does not generally survive scrutiny in multilocation economies. The reason is that, even in the presence of only a third location C, if this is large and well connected to both A and B, an increase in A's expenditure share may well map into a less than proportionate increase in its output share as location C drains away some firms. In more extreme cases, an increase in the expenditure share of a location may even lead to a decrease in its output share. This would be the case, for instance, if location C were a transport hub or gate. A hub is a location with better accessibility to all other locations; a gate is a location through which most goods flow in and out of a country. Positive demand shocks to any other location could result in supply expanding in the hub or in the gate and contracting elsewhere. Hence, agglomeration is more likely to take place in the presence of and close to hubs and gates. By analogy, preferential trade liberalization attracts firms to liberalizing countries while repelling them from other countries. The reason is that liberalizing countries improve their accessibility and become better export bases: they gain better access to one another's markets while maintaining the same ease of access to other countries' markets.

In contrast to the home-market effect, other insights of two-location models survive scrutiny for multilocation economies. First, if a location has a sufficiently large share of the economy's expenditures, a single cluster will emerge and will be hosted by that location (dominant market effect). Second, starting with prohibitive trade barriers, freer trade initially leads to a more uneven spatial distribution of firms (magnification effect).

Finally, having more than two locations allows one to model complex economies in which countries are themselves divided in regions. In this more complex scenario, one has to distinguish between international and interregional trade barriers while also keeping in mind that labor mobility is negligible at the international level but much greater at the regional level. The result is that agglomeration within countries is shaped mainly by interregional trade impediments while agglomeration between countries is shaped mainly by international trade barriers. This result has been derived both with and without interregional migration as well as with and without input-output linkages.

To summmarize, by relating the emergence and the strength of localized externalities to a set of well-defined microeconomic parameters, NEG allows one to understand the effects of international trade, migration, and capital flows on the geographical patterns of economic development in the world economy.

See also capital accumulation in open economies; fragmentation; linkages, backward and forward; location theory; monopolistic competition; New Trade Theory; oligopoly; technology spillovers

FURTHER READING

Baldwin, Richard, Rikard Forslid, Philippe Martin, Gian marco Ottaviano, and Frederic Robert Nicoud. 2003. *Economic Geography and Public Policy*. Princeton, NJ: Princeton University Press. A thorough analysis of a wide range of NEG models and a systematic investigation of how they affect our thinking about welfare and public policy issues.

Brakman, Steven, Harry Garretsen, and Charles van Mar rewijk. 2001. *An Introduction to Geographical Economics*. Cambridge: Cambridge University Press. A simplified presentation of NEG models for advanced undergrad uate and first year graduate students.

Fujita, Masahisa, Paul Krugman, and Anthony Venables. 1999. *The Spatial Economy: Cities, Regions, and Inter national Trade*. Cambridge, MA: MIT Press. The first book and still the main reference on NEG.

Fujita, Masahisa, and Jacques François Thisse. 2002. *Eco nomics of Agglomeration: Cities, Industrial Location, and Regional Growth*. Cambridge: Cambridge University Press. A rigorous synthesis of spatial economics from traditional location theory to models of systems of cities, from microfounded urban structure to NEG.

Henderson, Vernon, and Jacques François Thisse, eds. 2004. *Handbook of Regional and Urban Economics*.

Vol. 4. Amsterdam: Elsevier. A presentation of the most important recent developments in spatial economics in terms of both theory and empirics.

GIANMARCO I. P. OTTAVIANO

■ New Open Economy Macroeconomics

The New Open Economy Macroeconomics (NOEM), also referred to as new Keynesian open economy models, is a class of models that addresses open economy issues in a framework with dynamically optimizing agents with nominal rigidities (prices or wages that change slowly in response to shocks) and market imperfections (firms or households with some ability to set their own prices or wages). The NOEM model has become a "workhorse" framework for studying a variety of questions in international macroeconomics. These questions include whether international policy coordination could increase the welfare of households by improving economic outcomes, the welfare implications of simple monetary policy rules in an open economy, and how exchange rate behavior is affected by how firms set prices. As NOEM is more a framework than a particular model and one that economists use as the basis of current investigations, the assumptions of the model are still evolving. What follows is a short description of the original NOEM model written by the economists Maurice Obstfeld and Kenneth Rogoff (see Obstfeld and Rogoff 1996, 2000; Obstfeld 2001), some of its key results, and then several extensions of the model.

The Baseline NOEM Model Assume that in two symmetric countries, home and foreign, each household produces a differentiated good. Households produce goods using only their own labor but require increasing amounts of labor to produce additional goods (i.e., the production function exhibits decreasing returns to scale). Each product is distinct and is produced only by a particular household, such that each household has some market power in other words, some scope to choose the price at which it sells its good. Knowing what the demand for its good will

be at every price, each household chooses the price of its good in its own currency (called "producer-currency pricing," PCP), and during the following period the good's price will be fixed. Admittedly, this last assumption is somewhat ad hoc, but it is meant to capture the empirical fact that prices for most goods are not updated continuously, perhaps because there are some costs associated with changing prices. Since the price is set above the household's marginal cost, in the short run it is in the household's interest to supply more goods at the fixed price in response to small fluctuations. Both the law of one price (each foreign good costs its price set in the foreign currency converted to home currency by the nominal exchange rate) and purchasing power parity (which in one form says that the foreign bundle of goods costs the same as the home bundle of goods when converted into the home currency) hold. Note that this setup is equivalent to a profit-maximizing firm setting prices and wages paid to households, and the households supplying their labor to the firm as demanded.

Home and foreign households' utility depends on their consumption and work: they are made happier the more they consume and less happy the more they have to work. In particular, households would like to consume some of each of the goods produced in both the home and foreign economies. Households face budget constraints, however. Their income from selling their products, accumulation or decumulation of assets (which include both money and internationally traded bonds), and transfers from the government must equal their spending on consumption and the taxes they pay. This budget constraint must be satisfied across time, so home households could borrow from foreign households to purchase goods during a particular period but would have to pay foreign households back with interest. Each household chooses how much to produce in order to maximize its lifetime utility subject to its budget constraint.

This model will have some of the same effects found in the Mundell-Fleming model. For example, monetary policy, in the form of an unexpected

increase in the money supply in the home country, will lead to households in the home country increasing their demand for all consumption goods both home and foreign. It also will lead to a depreciation of the exchange rate, which temporarily raises domestic income relative to foreign income, and results in a current account surplus for the home country. This current account surplus increases the home country's net domestic claims on foreigners. All of these effects are largely permanent. Households in the home country will increase not only how much they consume today but also how much they consume in all future periods, because they will use some of their current income increase to acquire more assets. Although the real interest rate may move to offset some of this effect, changes to relative consumption demand will be permanent because domestic and foreign households face the same real interest rate.

Contrast with Earlier Models This baseline NOEM model is different in several key ways from both the Mundell-Fleming model and the intertemporal approach to the current account that preceded it. Unlike the intertemporal approach to the current account (earlier multiperiod optimizing models of the open economy), the NOEM model includes nominal rigidities and market imperfections. These features allow the NOEM model to look at the short-run effects of policy decisions, including any effects of monetary policy. Unlike the Mundell-Fleming model, this NOEM model is a dynamic general equilibrium model: it has explicit microfoundations (all agents are given objective functions to determine their behavior) and it has intertemporal choice (agents are making choices taking into account future periods). These differences allow the models to look at the *dynamic* effects of policies or shocks (as opposed to solely comparative *statics*), in particular on the current account and budget deficits. In addition, the NOEM model can evaluate different policies based on explicit welfare criteria that are consistent with the structure of the model.

For example, unlike later versions of the Mundell-Fleming model, in reaction to the home central bank loosening monetary policy this version of the NOEM model has no exchange rate overshooting,

where the exchange rate moves by a greater amount immediately than it will in the long run. This result is of interest to international economists because they would like to explain why exchange rates seem to move more than can be explained by changes in other macroeconomic variables or in ways contrary to what their models would suggest. For the reasons described earlier, home consumption relative to foreign consumption will immediately move to its new level, as will the relative money supplies in the two countries, and therefore the exchange rate.

Moreover, in the Mundell-Fleming model, the unexpected easing of home-country monetary policy raises output in the home country and lowers it in the foreign country, as the depreciation of the exchange rate makes home goods cheaper for foreigners. The depreciation is therefore said to have a beggar-thy-neighbor effect. In contrast, in the original version of the NOEM model, home and foreign households benefit equally from the easing of monetary policy in the home country. Both gain because households in each country have some monopoly power over their output and produce less individually than they would if society as a whole would choose their output for them. Monetary policy increases the utility of all households because it causes the expansion of output, and because both home and foreign households are equally affected by the market distortion. The point, however, is not so much the differences in results extensions to the original model can restore some of the conclusions from earlier models but rather the ability to look at dynamics and welfare conclusions.

Extensions of the Original NOEM Model Although the quantity of published research makes it difficult to discuss all of the extensions, some extensions of the original model have changed its basic results sharply. For example, the exchange rate will "overshoot" in response to a monetary policy shock if the two countries are no longer symmetric. If households in each country have "home bias" that is, prefer their own country's goods more than the other's goods or if some of the goods that each produces are not traded (such as most houses and some services), then the real exchange rate (the exchange rate adjusted for differences in each country's

inflation) will move in response to shifts in wealth between the two countries and differences in the real interest rate in the home and foreign nations. Because the tastes of the two households are no longer identical, purchasing power parity will no longer hold.

Extensions to the model can also change its welfare implications. Differences in the symmetry of home and foreign households can result in shocks, such as the home monetary policy shock, having a larger effect on the welfare of one country. With nontraded goods or home bias, households in the home country will gain more when domestic output increases. The degree to which households would be willing to substitute between goods (also called the elasticity of substitution), in other words, consume alternative goods, either another home good or a foreign good, can affect how they benefit from (or are harmed by) shocks. Therefore, if households are more willing to substitute for goods from the other country than they are for goods from the same country, then the unexpected home monetary policy expansion will benefit the home country more than the foreign one. If the difference between the two elasticities of substitution is large enough, then the welfare of the foreign country will be reduced a different version of the beggar-thy-neighbor effect from earlier models. If instead households are more willing to substitute for a good from the same country than for a good from the other country, perhaps because each country specializes in a certain type of goods, the foreign country will gain more: the home country has to work more but cannot consume as much because of the strong adverse movement in the cost of its imports relative to its exports, its terms of trade. The policy may even have a "beggar-thyself" effect if the difference in elasticities is large enough.

One extension that has been particularly discussed in the literature is whether agents use PCP or price their goods in the currency of the other country, referred to as "local-currency pricing" (LCP). Assuming LCP is attractive because it helps the model explain several things we observe in the real world: the nominal and real exchange rate will be perfectly correlated in the short run; changes in exchange rates will have little or no short-term pass-through to consumer prices; and because the exchange rate has to move more to affect the relative import price, the nominal exchange rate will be more volatile. Assuming LCP, however, means that, contrary to empirical evidence, a country's terms of trade will improve when its exchange rate depreciates (Obstfeld 2001). Several papers try to have agents optimally choose which currency to price in; this choice can depend on how easily people in both countries can share risk, how variable the exchange rate is, and the size of a firm's market share in the other country.

Neither PCP nor LCP makes the distinction between consumer prices and wholesale or intermediate prices, and as a result neither can capture both the relatively low pass-through of exchange rate movements to consumer prices and the larger pass-through to wholesale import prices. A number of authors explore this distinction by extending the model to include marketing and a distribution network goods require transportation and retail services, which are two nontraded goods, in order to be sold to households or to have nontraded goods produced by combining imported and domestically produced intermediate goods. Moreover, including a distribution sector in the model can help explain the behavior of the real exchange rate relative to movements in other macroeconomic variables, such as consumption.

In our description of the original NOEM model, all agents had perfect foresight and shocks were completely unexpected; in other words, they thought they knew everything that was going to happen in the model, and that anything could change came as a complete surprise to them. A number of researchers have extended NOEM models to a stochastic framework in which agents know that unexpected events will hit the economy and may even know the distribution of these shocks, but they do not know in any given period what the shock will be. In a stochastic model, where some prices or wages are set in advance of shocks, uncertainty or the volatility of variables can affect the means of variables in the economy, because households do not like risk. So risk can affect things such as welfare, the exchange rate, terms of trade, consumption, and how prices are set.

For example, since workers like to consume more but dislike working, if households are able to consume a large amount at the same time they are required to work more, they may wish to have higher wages to reduce the amount of work they have to do.

To take the example of monetary policy again, with floating exchange rates, more variable monetary policy (with no other shocks) can lead to a more variable nominal exchange rate and higher preset wages. These changes can increase the volatility of household consumption and move production in the economy farther away from the optimal level, and as a result lower welfare in the economy. One question that has particularly interested international economists is whether exchange rate volatility reduced the flow of goods, services, and assets traded in the world. The economists Philippe Bacchetta and Eric van Wincoop (2000) show that nominal exchange rate volatility does not necessarily have a negative effect on trade flows, although it probably has a negative effect on capital flows. Despite this theoretical work that shows a link between exchange rate uncertainty and both prices and real macroeconomic variables, however, earlier empirical results are mixed.

Researchers have also used the NOEM framework, with its explicit specification of a welfare criterion, to evaluate a variety of simple monetary policy rules and to solve for "optimal" monetary policy rules in an open economy. Whether a monetary policy rule, an equation for the central bank's policy instrument, includes some form of the exchange rate depends in part on the degree of pass-through from exchange rate to consumer prices. When firms price their goods only in their own currency, the home country's monetary policy may affect only the home country's producer currency prices since there is complete pass-through from exchange rate movements; foreign producer prices denominated in foreign currency will not change. In this case, the central bank should attempt to stabilize only domestic prices. With incomplete pass-though firms may charge higher export prices to compensate them for exchange rate risk. Because the central bank can affect the variability of the exchange rate and there-

fore possibly lower the risk premium charged on imports, it should no longer target the domestic price level only. Because policy decisions in one country can have implications for welfare in other countries, economists have also looked at whether policy coordination among policymakers in different countries yields welfare benefits.

In addition to the extensions outlined here, researchers have tried to address a variety of other questions, including currency crises, the short-run sustainability of the current account, the persistence of changes in the exchange rate, and the implications of agents not being able to completely share risk across markets. Policymakers use these models as well. For example, in recent years central banks in a number of countries as well as international organizations have built empirically relevant NOEM models to forecast the economy and to analyze alternative scenarios.

See also balance of payments; beggar-thy-neighbor policies; currency crisis; exchange rate pass-through; exchange rate regimes; exchange rate volatility; home country bias; international policy coordination; international reserves; monetary policy rules; Mundell-Fleming model; nontraded goods; purchasing power parity; real exchange rate

FURTHER READING

Bacchetta, Philippe, and Eric van Wincoop. 2000. "Does Exchange Rate Stability Increase Trade and Welfare?" *American Economic Review* 90 (December): 1093 1109. One of the earliest papers with a stochastic version of the NOEM framework.

Bowman, David, and Brian Doyle. 2003. "New Keynesian, Open Economy Models and their Implications for Monetary Policy." In *Price Adjustment and Monetary Policy*. Ottawa: Bank of Canada, 247 93. A longer and more in depth survey piece on NOEM models and ex tensions to them with a particular emphasis on issues related to monetary policy.

Corsetti, Giancarlo, Luca Dedola, and Sylvain Leduc. 2005. "DSGE Models of High Exchange Rate Volatility and Low Pass Through." CEPR Discussion Paper 5377 (December). London: Center for Economic Policy Re search. Available at http://www.cepr.org/Pubs/new dps/

dplist.asp?dpno=5377. Incorporates a distribution sector in the NOEM framework and tries to explain empirical observations about exchange rate behavior.

Corsetti, Giancarlo, and Paolo Pesenti. 2005. "The Simple Geometry of Transmission and Stabilization in Closed and Open Economies." NBER Working Paper 11341. Cambridge, MA: National Bureau of Economic Research. Available at http://www.nber.org/papers/w11341/pdf. Provides an excellent and relatively accessible discussion of optimal monetary policy, as well as an overview of the NOEM model and pass through.

Devereux, Michael, Charles Engel, and Peter Storgaard. 2004. "Endogenous Exchange Rate Pass Through When Nominal Prices Are Set in Advance." *Journal of International Economics* 63: 263 91. A recent example of trying to model firms' choice over which currency to price their exports.

Erceg, Christopher, Luca Guerrieri, and Christopher Gust. 2006. "SIGMA: A New Open Economy Model for Policy Analysis." *International Journal of Central Banking* 4 (March): 1 50. Available at http://www.ijcb.org/journal/ijcb06q1a1.htm. Outlines one of many open economy models being used for policy analysis at international organizations and central banks around the world.

Obstfeld, Maurice. 2001. "International Macroeconomics: Beyond the Mundell Fleming Model." *IMF Staff Papers* 47 (Special Issue): 1 39. Puts the New Open Economy Macroeconomics approach in the context of the history of international macroeconomics from the classical paradigm of the late 19th century and early 20th century through today.

Obstfeld, Maurice, and Kenneth Rogoff. 1996. *Foundations of International Macroeconomics.* Cambridge: MIT Press. Chapter 10 provides a textbook introduction to the earliest version of the NOEM model, as well as chapters on the Mundell Fleming and intertemporal approaches.

———. 2000. "New Directions for Stochastic Open Economy Models." *Journal of International Economics* 50 (February): 117 53. Outlines a benchmark stochastic version of standard model.

Svensson, Lars, and S. van Wijnbergen. 1989. "Excess Capacity, Monopolistic Competition, and International Transmission of Monetary Disturbances." *Economic Journal* 99 (September): 785 805. The most cited immediate precursor to the original Obstfeld and Rogoff NOEM paper.

BRIAN DOYLE

■ **New Trade Theory**

New Trade Theory is the descriptive term for theories that assume imperfect competition and increasing returns to scale (internal scale economies at the firm level) in order to explain international trade. In contrast, traditional trade theories, such as the Ricardian or the Heckscher-Ohlin model, assume perfect competition and constant returns to scale, and attribute the emergence of international trade to cross-country opportunity cost differences (i.e., comparative cost advantage) caused by either differences in production technology (Ricardian model) or factor endowments (Heckscher-Ohlin model).

Reasons for Developing New Trade Theory
The New Trade Theory originated in the late 1970s and early 1980s when advances in the industrial organization literature were incorporated into trade theory. It was developed to explain four major stylized facts (Helpman and Krugman 1985):

Fact 1: *The increase in the trade to income ratio, and the high and increasing percentage of trade taking place between industrialized nations, even though these countries are seemingly becoming more similar.*

Traditional models predict that trade benefits increase the more different trading partners are. This is puzzling because, in reality, trade is much more prevalent between quite similar countries, in particular between industrialized nations. For example, in 2005, the United States exported (imported) $186.4 ($308.8) billion worth of goods and services to (from) the European Union. In the same period, the U.S. export (import) value to (from) Africa was only $15.5 ($65.2) billion (Census FT900 Report).

Fact 2: *The high percentage of intraindustry trade, particularly between industrialized nations.*

Traditional trade theories cannot explain the existence of intraindustry trade. In the Ricardian and the

Heckscher-Ohlin models, only interindustry trade takes place. That is, if one country exports cars to another country, it does not import cars from the same country. In contrast, trade statistics for industrialized nations consistently reveal intraindustry trade to account for well above 50 percent of total trade. Although part of this may just be an aggregation problem, that is, different products are aggregated into one industry so that actual interindustry trade appears as intraindustry trade, it is clearly not the whole story.

Fact 3: *Intrafirm transactions and foreign direct investment by multinational firms.*

In the constant returns perfect competition world of the traditional theories, it is essentially impossible to talk about the scope of activities carried out within firms because there is no true concept of "a firm," the number and size of firms being indeterminate.

Fact 4: *Trade liberalization and the associated increase in factor productivity in spite of only limited resource reallocation.*

According to the traditional theories, trade liberalization results in substantial factor reallocation, most notably in the Ricardian model, where one industry closes down completely and all resources concentrate in the industry for which a country has a comparative advantage. The effect is qualitatively similar in the Heckscher-Ohlin framework, except that specialization is incomplete. In reality, however, trade liberalization seems to cause rather limited resource reallocation.

Some, usually nonacademic, literature claims that the raison d'être of the New Trade Theory is to provide arguments against free trade. Although there is research showing that active trade policy may improve welfare under imperfect competition, these findings are usually too assumption-dependent to warrant a recommendation for a widespread deviation from free trade.

Models in the New Trade Theory can be distinguished by the type of imperfect competition. The first approach uses the monopolistic competition model with differentiated goods, typically as proposed by Dixit and Stiglitz (1977). Research in this area began with seminal articles by Krugman (1979,

1980, 1981). The second approach, starting with Brander (1981) and Brander and Krugman (1983), uses an oligopolistic setup.

Explaining Intraindustry Trade The basic model (Krugman 1979) assumes that consumers value variety, but it is also possible to set up the model so that each consumer has a preferred variety that differs across consumers (Helpman and Krugman 1985). Each variety is produced by a single firm using labor. The cost function of the firm consists of fixed and variable costs; marginal cost is constant, and average cost is decreasing (internal economies of scale). The consumers' utility function is such that the price elasticity of demand for any variety is finite and exceeds one, so that the profit-maximizing price exceeds marginal cost by a positive markup. The number of firms is limited by the zero profit condition.

When the domestic country opens up to trade with an otherwise similar economy, foreign consumers create additional demand for domestic varieties. In equilibrium, individual consumption per variety and the price-wage ratio both fall. The trade-induced increase in demand allows firms to move down their average cost curve and increase production. At the same time, the full employment condition still holds. Hence, since firm-level production increases, the number of domestic firms goes down. Yet the number of varieties, domestic plus foreign, available to domestic consumers increases.

Gains from trade arise for two reasons. First, trade allows firms to better exploit economies of scale; as a result, the domestic real wage rises. Second, domestic consumers have access to a larger number of varieties.

To analyze the impact of transportation costs and home market effects, consider two differentiated industries *A* and *B* and two groups of consumers in countries *1* and *2* (Krugman 1980). The first group consumes only products from industry *A* and the second group only products from industry *B*. Type *A* consumers have a higher population share in country *1* and type *B* consumers in country *2*, while country population levels are the same. Due to positive transportation costs, the consumption ratio of foreign to domestic goods is below one in both coun-

tries. In this case, each country will produce relatively more of the product for which it has the larger home market. If countries' tastes are sufficiently dissimilar (i.e., if the population shares of A-consumers in 1 and B-consumers in 2 are high), specialization will be complete: industry A will be located only in country 1 and industry B only in country 2. The model thus explains how a bigger home market can give an industry a comparative advantage.

To investigate the determinants of intraindustry trade shares, consider two differentiated-goods industries and two sector-specific inputs (Krugman 1981). To keep things simple, let the two countries have equal population size, but country 1 has a share z < 50 percent of industry B specific labor and country 2 has a share z of industry A specific labor. Whereas the trade volume is constant and does not depend on z, the share of intraindustry trade becomes maximal as z approaches 50 percent. This model can thus explain why we see more intraindustry trade between countries that are more similar in terms of factor endowments.

Intraindustry trade also arises in a model with one domestic and one foreign market and two firms, domestic and foreign, who produce a homogeneous good and compete Cournot style (i.e., each firm treats the output level of its competitor as fixed) in both markets (Brander 1981). The markets are segmented, and shipping goods from one country to the other is costly. In this case, as long as the per unit transportation cost does not exceed the monopoly price markup over marginal cost, both firms will sell goods in both markets, so intraindustry trade arises.

Topics in New Trade Theory Traditional trade theory shows that small countries (those that cannot influence world market prices) typically have nothing to gain from pursuing an active trade policy. This seems at odds with the empirical observation that trade intervention is widespread across countries and industries. Although both monopolistic competition and oligopoly models can provide additional reasons for active trade policy (see Helpman and Krugman 1989; Rivera-Batiz and Oliva 2003), the literature on what is known as strategic trade policy has focused more on oligopoly. Using active trade policy, the domestic government can alter the strategic position of domestic firms against foreign competitors and help them capture additional rents. Oligopoly models can also be used to explain why a government may prefer tariffs over quotas or other nontariff barriers and vice versa; in contrast, tariffs and quotas are usually equivalent in the perfect competition model. Although imperfect competition can be used to show why active trade policy may make sense, this does not imply that free trade should be abandoned. In the absence of correct information about market structure and cost, governments can easily lower welfare by using active trade policy. Recent research, using asymmetric information models, has investigated whether governments can induce firms to reveal this information.

Models of imperfect competition can also be employed to explain multinational corporations, outsourcing, and the exporting decision of firms. Summarizing Dunning (1988), multinational corporations arise due to three reasons (OLI theory):

1. *Ownership advantage*: Knowledge capital (headquarter services) can be used at the same time by different production facilities.
2. *Location advantage*: Having a subsidiary in a foreign market makes sense when the market is large and trade costs are high (for horizontal firms) or when trade costs are low and factor prices differ substantially across countries (for vertical firms).
3. *Internalization advantage*: Transfer of knowledge-based assets in order to have other firms produce inputs for one's own firm is risky.

This last point leads to the question of firm organization; namely, when would firm A want to build a subsidiary instead of outsourcing production to another firm B? The answer depends on transaction costs. Letting firm B produce the required input may be more efficient in terms of production costs, but problems may arise because of input specificity and the related holdup problem. Suppose the input that firm B manufactures can be used only by firm A. Assuming incomplete contracts, firm A can then

force firm *B* to sell below production value (holdup problem). Foreseeing this, firm *B* may be tempted to produce an input that is not well suited for *A* but can be more easily sold to other firms. Grossman and Helpman (2002) and Antras (2003), among others, have integrated the transaction cost approach into a monopolistic competition trade model to investigate when outsourcing will occur.

In reality, only a small fraction of firms actually export. The question of which firms will export can be addressed in a monopolistic competition model, where different firms have different variable costs and face per unit transportation costs and fixed costs of exporting (Melitz 2003). In this framework, it can be shown that when a country opens up to trade, less productive firms will exit the market and only a more productive subset of the remaining firms will decide to export.

New Growth Theory (Grossman and Helpman 1991) and New Economic Geography (Fujita, Krugman, and Venables 1999) are closely related to the New Trade Theory since they also heavily rely on the Dixit-Stiglitz model of monopolistic competition and emphasize the importance of trade. These theories are usually not considered part of the New Trade Theory, however, and are thus not discussed here any further.

The New Trade Theory, based on the concepts of imperfect competition and increasing returns to scale, complements traditional trade theories and has contributed substantially to a better understanding of the world economy and its patterns of trade.

See also comparative advantage; economies of scale; foreign direct investment (FDI); foreign direct investment: the OLI framework; gains from trade; Heckscher-Ohlin model; intraindustry trade; monopolistic competition; New Economic Geography; oligopoly; Ricardian model

FURTHER READING

Antras, Pol. 2003. "Firms, Contracts, and Trade Structure." *Quarterly Journal of Economics* 118 (4): 1375 1418. Integrates incomplete contracting model of firm boundaries into monopolistic competition trade model.

Brander, James A. 1981. "Intraindustry Trade in Identical Commodities." *Journal of International Economics* 11 (1): 1 14. Shows how intraindustry trade in homoge neous goods can arise in segmented oligopoly markets.

Brander, James A., and Paul R. Krugman. 1983. "A 'Re ciprocal Dumping Model' of International Trade." *Journal of International Economics* 15 (3 4): 313 21. Discusses price discrimination and reciprocal dumping in segmented oligopoly markets.

Dixit, Avinash K., and Joseph E. Stiglitz. 1977. "Mono polistic Competition and Optimum Product Diversity." *American Economic Review* 67 (3): 297 308. Develops model of monopolistic competition and differentiated goods.

Dunning, John H. 1988. *Explaining International Produc tion*. London: Unwin Hyman. Lays out the eclectic paradigm, also known as OLI theory, for explaining multinational enterprises.

Fujita, Masahisa, Paul Krugman, and Anthony J. Venables. 1999. *The Spatial Economy*. Cambridge, MA: MIT Press. A book length treatment of the New Economic Geogra phy, which analyzes the causes for spatial economic ag glomeration in a general equilibrium framework.

Grossman, Gene M., and Elhanan Helpman. 1991. *In novation and Growth in the Global Economy*. Cambridge, MA: MIT Press. A book length treatment of the New Growth Theory, which considers intentional techno logical innovation as the driving force behind economic growth.

———. 2002. "Integration versus Outsourcing in Industry Equilibrium." *Quarterly Journal of Economics* 117 (1): 85 120. Discusses firms' integration versus outsourcing choice in transaction cost setting.

Helpman, Elhanan, and Paul R. Krugman. 1985. *Market Structure and Foreign Trade*. Cambridge, MA: MIT Press. Combines Heckscher Ohlin model with mo nopolistic competition and differentiated goods model and discusses the impact on trade volume and gains from trade.

———. 1989. *Trade Policy and Market Structure*. Cam bridge, MA: MIT Press. Discusses impact of imperfect competition on optimal trade policy.

Krugman, Paul R. 1979. "Increasing Returns, Monopolistic Competition, and International Trade." *Journal of In*

ternational Economics 9 (4): 469 79. Shows gains from trade in monopolistic competition model.

. 1980. "Scale Economies, Product Differentiation, and the Pattern of Trade." *American Economic Review* 70 (5): 950 59. Introduces transportation costs and home market effects into monopolistic competition model.

. 1981. "Intraindustry Specialization and the Gains from Trade." *Journal of Political Economy* 89 (5): 959 73. Discusses determinants of intra vs. interindustry trade in monopolistic competition model.

Markusen, James R. 2002. *Multinational Firms and the Theory of International Trade.* Cambridge, MA: MIT Press. A book length treatment of the theory of multi national firms under oligopoly and monopolistic com petition.

Melitz, Marc J. 2003. "The Impact of Trade on Intra Industry Reallocations and Aggregate Industry Pro ductivity." *Econometrica* 71 (6): 1695 1725. Provides a model of firms' exporting choice in a monopolistic competition framework.

Rivera Batiz, Luis A., and Maria Angels Oliva. 2003. *In ternational Trade.* Oxford: Oxford University Press. A comprehensive textbook on international trade with large part devoted to new literature, in particular on trade policy.

XENIA MATSCHKE

■ **nominal anchor**

See discipline

■ **nondiscrimination**

The nondiscrimination obligation commits World Trade Organization (WTO) members to avoid dis criminating among products solely on the basis of their national origin. This obligation is manifested in two basic provisions in the WTO. The most-favored-nation (MFN) clause essentially requires that equal treatment be afforded to all imported goods, irre spective of their origin, as long as they are "like"; the basic incarnation of MFN is found in Article I of the General Agreement on Tariffs and Trade (GATT). The other nondiscrimination principle is the na tional *treatment* (NT) *obligation, w*hich appears in its original form as Article III in the GATT. It requires WTO members to treat imported goods no less favorably than domestically produced "like" prod ucts. By virtue of the MFN obligation, a WTO member cannot treat goods originating in a non-WTO member better than a like good originating in a WTO member. The MFN obligation applies not only to trade instruments (tariffs), but also, by virtue of Article I.1 GATT, to all *domestic* measures that affect trade. NT, on the other hand, covers domestic instruments only. Both MFN and NT are relevant irrespective of whether tariff commitments have been entered on a particular good. Thus, with respect to tariffs, for example, MFN is relevant both for the tariff levels that members agree not to exceed the "bound rates" and for the tariff levels that are ac tually applied (which may be lower).

The MFN principle has a long history. For in stance, Hudec (1988) reports that the medieval city of Mantua (Italy) obtained from the Holy Roman emperor the promise that it would always benefit from any privilege granted by the emperor to "what-soever other town." According to Jackson (1997, 158), the term *MFN* appears for the first time at the end of the 17th century. During the 19th century, the provision appeared in a number of treaties across European states. For instance, the Cobden-Chevalier Treaty of 1860, liberalizing trade between Great Britain and France, included an MFN clause guaran teeing that a signatory would not be treated worse than any other state with which the other signatory had, or would assume, trade relations. Such schemes, however, did not amount to a worldwide nondis criminatory trade; if at all, nondiscriminatory trade existed between a subset of all states, those that had entered into a similar contractual arrangement es pousing the MFN clause.

Nondiscrimination provisions appear in the WTO Agreement in all three Annex 1 Agreements, that is, in the General Agreement on Tariffs and Trade (GATT), in the General Agreement on Trade

in Services (GATS), and in the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPs). The GATS and the TRIPs contain MFN and NT provisions that are essentially identical to those in GATT, as described above. However, the GATS NT provision (Article XVII) covers only domestic instruments with respect to sectors that a WTO member has decided to liberalize, and to the extent that no liberalization commitment has been entered, it is of no relevance. We will focus on the MFN and NT obligations in the GATT context. There are also NT-like provisions in some of the other agreements constituting the WTO Agreement. For instance, the "consistency" requirement embedded in Article 5.5 of the Agreement on the Application of Sanitary and Phytosanitary Measures (SPS) has a strong flavor of NT by requiring that WTO members treat comparable risks associated with imported and domestic products in similar fashion.

WTO adjudicating bodies have understood MFN and NT to ban both de jure and de facto discrimination. The former concept covers cases where a discriminatory treatment is afforded by virtue of the origin of the product; the latter refers to cases where treatment that is nondiscriminatory on its face effectively confers an advantage to the domestic good. While the de jure cases are easy to delineate, this is not the case with de facto violations, where case law has developed in a rather erratic manner.

Nondiscrimination and Negative Integration
The inclusion of the nondiscrimination obligations is largely the direct result of the decision to construct the GATT as a negative integration contract: policies are defined unilaterally and, to the extent that there are international spillovers, they will be internalized by virtue of the nondiscrimination obligation. This means that a WTO member can in practice adopt any regulation (rational or irrational) to the extent that it does not discriminate. The agreements on Technical Barriers to Trade (TBT) and Sanitary and Phytosanitary measures (SPS) do request a certain degree of rationality from regulatory interventions, but without negating the nondiscrimination discipline. For example, a WTO member has (at least in principle) to base its SPS regulations on scientific evidence. Regulation not based on scientific evidence (when such evidence is available) is SPS-inconsistent, even if nondiscriminatory.

A widespread view among policymakers, lawyers, and many economists is that there are a number of strong economic rationales for nondiscrimination. For instance, it seems to be commonly believed (in particular among noneconomists) that nonuniform tariff structures give rise to inefficient production and consumption patterns in a static sense. Other arguments in favor of nondiscrimination hold that it eases tariff negotiations, or may prevent the formation of preferential trading agreements that are formed to exploit market power in world markets. However, a general theoretical prima facie case for nondiscrimination is not as easily advanced as might be thought. Indeed, Johnson (1976, 18) goes as far as arguing that "the principle of non-discrimination has no basis whatsoever in the theoretical argument for the benefits of a liberal international trade order in general, or in any rational economic theory of the bargaining process in particular."

Nondiscrimination in the Form of Most-Favored-Nation Treatment The MFN obligation, as interpreted in case law, consists of the following elements:

1. with respect to, in principle, all measures that affect trade either de jure or de facto, any advantage granted to goods originating anywhere in the world;
2. must be extended to the like products;
3. originating in any WTO member;
4. immediately and unconditionally.

It emerges from the case law that two products sharing the same tariff classification will be deemed *like*. There is no WTO agreement on rules of origin, and, consequently, every WTO member can unilaterally define its own policy in this respect and apply it in a nondiscriminatory manner. *Immediately* means that the importing WTO member has to grant an advantage with no passage of time. *Unconditionally* means that it cannot impose conditions that it has not imposed to another beneficiary. The standard of review applied by WTO adjudicating

bodies in MFN cases is quite favorable to the complainant: there is no need to demonstrate intent to discriminate, and there is no need to demonstrate the resulting trade effects either.

There are two important exceptions to MFN: preferential trade agreements (PTAs), and special and differential treatment for products originating in developing countries that, by virtue of their origin, are granted in donors' markets better treatment than like products originating in developed countries.

A number of factors contribute to make an economic analysis of MFN complicated. First, in a world where free trade maximizes global welfare, there is of course no scope for tariffs at all, discriminatory or not. Any meaningful analysis of the desirability of MFN must therefore include a reason why there are tariffs at all. Consequently, it will involve comparisons of distorted equilibria, with associated second-best problems. Also, a study of MFN must involve at least three countries, with the plethora of different possible trade patterns and analytical difficulties this typically incurs.

Yet another difficulty in analyzing MFN is the fact that there is no general well-defined uniform structure with which to compare the nondiscriminatory tariff structure. This is problematic since it is a priori clear that the arbitrary choice concerning this level will have a profound effect on the outcome: if the uniform level is the same as the lowest discriminatory tariff, the uniform structure will most likely be preferred, while an equally nondiscriminatory structure, but set at the level of the highest discriminatory tariff, will most likely be worse from a welfare point of view. In neither case can the outcome be said to be related to discrimination, however. More generally, we lack a meaningful measure of the *degree* to which a structure fulfills MFN; it is not possible to simply "turn up" the degree of non-MFN and observe the outcome.

Despite these inherent complexities, several strands of economic theory can shed light on the impact of an MFN clause. Both the literature on optimal taxation and the industrial organization (IO) literature on price discrimination suggest reasons why discrimination may be socially desirable. For instance, if the raison d'être for tariffs is to raise government revenue, tariffs (like taxes in general) should be levied so as to minimize the resulting distortions, and this will often call for nonuniform structures.

Economic analyses of MFN can be divided into two categories. The first comprises models in which governments set tariffs unilaterally. In a typical setup, firms decide on investment, the level of which is influenced by firms' perceptions of tariff treatment, and thus on whether MFN has to be respected or not. A basic mechanism here is that MFN hinders ex post opportunistic taxation of economic rents and may thereby increase the ex ante private incentives for the creation of such rents. This mechanism lies behind several observations. For instance, by affecting the strategic interaction between firms and governments, an MFN clause may have a positive welfare impact even if the government would choose to set nondiscriminatory tariffs in its absence; and a government that, absent an MFN clause, would choose to discriminate may gain from being prevented from discriminating.

A second and more recent strand of literature is concerned with the role of MFN for multilateral trade liberalization, and in particular tariff negotiations. One fundamental role of trade agreements is to prevent negative externalities from nationally pursued trade policies. These international externalities may work through a number of different routes. For instance, they may take the form of changes in terms-of-trade, or through domestic prices affecting import demand. Bagwell and Staiger (2002) suggest that a central role of MFN is to channel these externalities through the terms of trade. This is important, since tariff negotiations can directly address terms-of-trade externalities, but are less effective to address other forms of externalities. Bagwell and Staiger also show how MFN may work in concert with other characteristic features of the GATT/WTO system, such as reciprocity, to make multilateral trade agreements immune to Article XXVIII GATT renegotiations.

The complexity of multicountry tariff negotiations is reflected in the wide variety of intuitively plausible but often contradictory arguments that have been advanced in the informal academic

literature and in policy discussions. For example, MFN is said to promote tariff liberalization by making trade agreements more credible; the increased cost of giving concessions makes it less attractive for a party to undermine an agreement by subsequently offering better terms of market access to a third country ("concession diversion"). MFN also makes it attractive for outsiders to enter into an existing agreement, since they get access to a package of low tariffs. And since entrants have to grant MFN, insiders get access to many foreign markets through the incentives for entry. On the other hand, MFN is also claimed to reduce the incentives to liberalize. It increases the costs of giving concessions, since the latter have to be given to all countries with which a country has MFN agreements; MFN makes large countries unwilling to make concessions to small countries, since in return for "peanuts" large countries have to extend their concessions to a large volume of trade; MFN reduces the benefit from a given concession since it has to be shared with other countries; MFN promotes free riding, since countries may opt to wait for agreements between other countries to spill over via MFN, rather than contribute with concessions themselves, and MFN also prevents countries from punishing free riding; or MFN prevents subsets of countries from going further in liberalization than what is desired by the rest of the world. A number of studies examine the validity of these types of claims. We will here just mention two of these.

The role of MFN to prevent concession diversion is at focus in Ethier (2004), who takes a very long-run perspective on its impact. Governments are assumed to initially form reciprocal bilateral agreements. To be meaningful, these must include MFN to avoid concession diversion. As more and more bilateral agreements are formed, the incentives to participate in further agreements gradually diminish, since each agreement has through the partner's MFN commitment to be shared by more and more other countries, and more and more market access has to be given away through a country's own MFN commitments. A process of liberalization through bilateral agreements will therefore eventually come to a halt. It will become necessary to internalize the ex-

ternal effects of any further agreements by making the agreements multilateral. Hence, MFN causes multilateralism, not the other way around. This study can also be seen as an illustration of the more general point that bilateral negotiations conducted under MFN generally are associated with externalities, since the outcome of such negotiations affect parties who are not present in the negotiations.

MFN may potentially cause free riding in at least two ways. One is that a country rejects an offer in order to let other countries reach agreements that it can benefit from without having to make concessions itself. This would be inefficient either because there would be delays in achieving an agreement, or because the agreement would feature higher tariffs compared to some other (undefined) situation. This possibility has as far as we know not found any support in the literature so far. For instance, Ludema (1991), which is one of the few studies that employ a noncooperative sequential bargaining model to study the impact of MFN on multilateral bargaining, shows how negotiators may find it optimal to devise equilibrium offers such that free riding does not occur, despite there being incentives and possibilities to free-ride.

As mentioned, in this entry we do not for the most part cover the literature on PTAs, since it tends not to focus on MFN as such. Nevertheless, some recent work on PTAs is of direct relevance to the issues at stake here. In particular, the literature has highlighted the interplay between the domestic political system and the formation of PTAs, as exemplified by the studies by Grossman and Helpman (1995), Levy (1997), and Krishna (1998), all of which suggest that MFN may have desirable welfare consequences by constraining the domestic political process.

Finally, there are some intuitively important aspects of MFN that have not been formally scrutinized, as far as we know. For instance, as noted already by Viner (1931), the administration of discriminatory tariffs is costly because of the need to keep track of product origin, and MFN thus significantly simplifies customs procedures. Another aspect that we believe is of considerable importance is the fact that MFN reduces the cost and complexity of

negotiations by reducing the number of possible bids and outcomes. Finally, under MFN countries may have incentives to use narrow product classifications in order to avoid having to extend concessions granted on an MFN basis. There should thus be reasons for countries to try to manipulate customs classification schemes.

To conclude, the implications of MFN for multiparty tariff negotiations are inherently complex. A large number of partial effects have been suggested, but the economic theory literature has only examined a few of these. It does support some of the claims concerning beneficial effects of MFN, but provides a far too scattered picture to serve as a basis for any more general claim concerning the desirability of MFN.

Finally, it is sometimes argued that MFN is today of limited practical importance, given the low average tariffs of developed countries on imports of industrial products. However, current tariffs are the result of a system built on MFN, and there is no guarantee a priori that the same levels could be supported in its absence. Also, there are important sectors such as agriculture, textiles, and services where barriers are still high and where MFN (or its absence) might clearly be important. Furthermore, as discussed above, the MFN principle does not apply only to tariff negotiations in the rounds, but also to many other facets of the WTO.

Nondiscrimination in the Form of National Treatment Article III GATT divides domestic policy interventions into fiscal (Article III.2) and nonfiscal (Article III.4) measures. With respect to fiscal measures, any tax differential across two "like" products will violate the NT discipline, whereas the tax differential across two "directly competitive or substitutable" (DCS) products must operate "so as to afford protection to domestic production," in order to be in violation of NT. Case law so far has held that large tax differentials across DCS products suffice for this requisite to be fulfilled. Two products are in a DCS relationship if they are considered by consumers to be interchangeable. A demonstration of interchangeability can be based on either econometric or noneconometric indicators. Two products are like if they are DCS *and* also share the same tariff

classification, to the extent that it is detailed enough. Turning to nonfiscal measures, a measure roughly speaking violates NT if it affords to an imported good less favorable treatment than that afforded to a domestic like product. The term *like* has here been interpreted in a manner that roughly corresponds to the DCS notion for fiscal measures.

Two domestic instruments are by virtue of Article III.8 GATT excluded from the coverage of NT: production subsidies and government procurement. But the GATT does not contain a list of domestic instruments that have to observe the NT discipline. WTO adjudicating bodies have adopted an all-encompassing attitude in this respect and have never so far refused the application of Article III GATT on the grounds that a particular measure does not come under its coverage. There is indirect legislative support for this approach: Article III.2 GATT stipulates that imported products shall not be subject "directly or indirectly" to higher taxation or levies than those imposed "directly or indirectly" to like domestic products. Hence fiscal measures that only indirectly hit products are covered by the NT discipline as well. Likewise, the term *affecting* in Article III.1, which includes nonfiscal as well as fiscal measures, is in the same broad vein; finally, the Interpretative Note to Article III states that measures enforced at the border should still be considered domestic, if they are designed to regulate both domestic and imported goods.

Case law has established that violations of the NT discipline can be justified through recourse to one or more of the public order grounds reflected in Article XX GATT. The intellectual merits of constructing Article XX as an exception to Article III are highly doubtful however, since both provisions contain a nondiscrimination clause. There is no doubt, though, that NT can lawfully be disrespected on national security grounds (Article XXI GATT).

Turning to the rationale for NT, the most striking feature of the economic literature is the near-complete nonexistence of studies of NT. The question of what role NT plays in trade agreement is part of the larger question of why trade agreements typically treat internal measures very differently from border

measures. Border measures are largely explicitly regulated; tariff levels are bound, import and export quotas as well as export subsidies are prohibited, and so forth. Internal measures, on the other hand, are left to be unilaterally determined by the contracting parties. Intuitively, the reason for this asymmetry seems to stem from a combination of two facts. First, a trade agreement is typically in place for a long period of time because of the costs of negotiating it, even if only binding border measures. Second, during the life span of the agreement various nonprotectionist regulatory needs are likely to arise. The agreement must therefore leave sufficient flexibility for the contracting parties to use domestic instruments for nonprotectionist reasons. It would be extremely costly to condition the agreement on all possible developments; on the other hand, it cannot leave internal instruments completely unregulated, since this would enable countries to use internal measures to undo whatever restrictions are agreed on regarding border measures. NT is the first line of defense against such behavior. That is, NT can be understood as an attempt to remedy problems caused by incompleteness of the agreement.

In order to save on contracting costs, an NT provision is likely to impose a less than perfectly flexible discipline on domestic measures. The question therefore arises as to whether the restriction increases welfare. Horn (2006) shows how an extreme version of an NT provision a dictum never to tax foreign products higher than domestic products may improve the contracting parties' welfare even if it would be desirable to tax foreign products higher from an efficiency point of view. Basically, the NT restriction makes internal instruments blunter tools for protectionism: domestic instruments can be used against foreign products, but they also have to hit domestic like products to the same extent. As a result, taxes will be used less to protect, and countries may go further in their tariff liberalization, creating overall gains from the imposition of the NT provision. Such a strict NT restriction will not completely eradicate the problem caused by incomplete contracting, however.

A fundamental weakness of the whole incomplete contract trade agreement literature is the fact that the structure of the incompleteness   what is contracted and what is not   is simply assumed. In practice, parties to an agreement can choose not only tariff levels, say, but also whether or not to include an NT provision. A natural question, therefore, is when, if at all, an NT provision is likely to be part of an optimal agreement. Horn, Maggi, and Staiger (2006) employ a model with explicit contracting costs to endogenously determine the incomplete structure of a trade agreement. It is shown how a strict NT-like provision of the above-mentioned type may be an optimal component of a trade agreement, by combining limited contracting costs with a degree of discipline on countries' use of domestic taxes for beggar-thy-neighbor purposes.

Overall, the analysis of the implication, and even more the optimal design, of NT has only just begun. This is an unfortunate state of affairs, considering the importance of NT and NT-like provisions in the WTO Agreement and in high-profile WTO trade disputes. Without NT (or a provision with a similar type of effect), agreements on border measures could be rendered entirely meaningless by opportunistic domestic policies; to take the simplest case, a domestic tax on imported products could perfectly replace a tariff that is bound through a trade agreement. Nonfiscal measures can be used in a similar nondiscriminatory fashion. It is thus not by chance that NT-like provisions appear in several of the main agreements regulating trade in goods, and also in regulation of trade in services and intellectual property in the WTO.

See also General Agreement on Tariffs and Trade (GATT); multilateralism; regionalism; World Trade Organization

FURTHER READING

Bagwell, Kyle, and Robert W. Staiger. 2002. *The Economics of the World Trading System*. Cambridge, MA: MIT Press. These authors have examined MFN in a series of writings, and this book summarizes some of this work.

Caplin, Andrew, and K. Krishna. 1988. "Tariffs and the Most Favored Nation Clause: A Game Theoretic Ap

proach." *Seoul Journal of Economics* 1: 267 89. One of the first formal economic analyses of MFN.

Ethier, Wilfred J. 2004. "Political Externalities, Non discrimination, and a Multilateral World." *Review of International Economics* 12 (3): 303 20. Discusses the role of MFN in the evolution of multilateralism.

Grossman, Gene, and Elhanan Helpman. 1995. "The Politics of Free Trade Agreements." *American Economic Review* 85 (September): 667 90. Highlights implica tions of MFN for the domestic political process.

Horn, Henrik. 2006. "National Treatment in Trade Agreements." *American Economic Review* 96 (1): 394 404. A formal economic analysis of implications of NT.

Horn, Henrik, Giovanni Maggi, and Robert W. Staiger. 2006. "Trade Agreements as Endogenously Incomplete Contracts." NBER Working Paper No. 12 745. Cam bridge, MA: National Bureau of Economic Research. Shows how NT can arise as an optimal part of a trade agreement in the presence of contracting costs.

Horn, Henrik, and Petros C. Mavroidis. 2001. "Legal and Economic Aspects of MFN." *European Journal of Poli tical Economy* 17 (2): 233 79. Summarizes core legal aspects of MFN and surveys the economic literature on the topic.

———. 2004. "Still Hazy after All These Years: The In terpretation of National Treatment in the GATT/WTO Case Law on Tax Discrimination." *European Journal of International Law* 15 (1): 39 69. Discusses the WTO dispute settlement case law on NT as it applies to taxa tion.

Hudec, Robert E. 1988. "Tiger, Tiger in the House: A Critical Evaluation of the Case against Discriminatory Trade Measures." In *The New GATT Round of Multi lateral Trade Negotiations: Legal and Economic Problems*, edited by Ernst Ulrich Petersmann and Meinhard Hilf. Deventer, Holland: Kluwer, 165ff. An early and notable legal view of discriminatory trade measures.

———. 1998. "GATT/WTO Constraints on National Regulation: Requiem for an 'Aims and Effect' Test." *International Lawyer* 32 (3): 619 45. Provides a so phisticated legal argument in favor of introducing an intent test in Article III.

Jackson, John H. 1997. *The World Trading System: Law and Policy of International Economic Relations*. Cambridge: Cambridge University Press. A comprehensive analysis of the WTO.

Johnson, Harry. 1976. "Trade Negotiations and the New International Monetary System." In *Commercial Policy Issues*, edited by G. Curzon and V. Curzon. Leiden: A. W. Sijthoff. An early economic analysis of trade ne gotiations.

Krishna, Pravin. 1998. "Regionalism and Multilateralism: A Political Economy Approach." *Quarterly Journal of Economics* 113 (1): 227 52. Highlights implications of MFN for the domestic political process.

Levy, Philip. 1997. "A Political Economic Analysis of Free Trade Agreements." *American Economic Review* 87 (4): 506 19. Highlights implications of MFN for the do mestic political process.

Ludema, Rodney. 1991. "International Trade Bargaining and the Most Favored Nation Clause." *Economics and Politics* 3 (1): 1 20. Uses noncooperative bargaining theory to highlight the implications of MFN.

Schwarz, Warren, and Alan O. Sykes. 1998. "The Positive Economics of the Most Favored Nation Obligation and Is Exceptions in the WTO/GATT System." In *Economic Dimensions in International Law*, edited by J. Bhandari and A. Sykes. Cambridge: Cambridge University Press, 43 75. A survey of the MFN literature, with particular emphasis on political economy aspects.

Staiger, Robert W. 1995. "International Rules and In stitutions for Trade Policy." In *Handbook of Interna tional Economics*, vol. 3, edited by G. Grossman and K. Rogoff. Amsterdam: North Holland, 1495 1551. A survey of the economic literature on trade agreements.

Viner, Jacob. 1924. "The Most Favored Nation Clause in American Commercial Treaties." *Journal of Political Economy* 32 (February). Reprinted in J. Viner. 1951. *International Economics*. Glencoe, IL: Free Press. Several of these arguments were made by Viner (1924, 1931). A very early discussion of economic and political impli cations of MFN.

———. 1931. "The Most Favored Nation Clause." Index VI (January). Stockholm: Svenska Handelsbanken. Reprinted in J. Viner. 1951. *International Economics*. Glencoe, IL: Free Press. See previous entry.

HENRIK HORN AND PETROS C. MAVROIDIS

■ nongovernmental organizations (NGOs)

Nongovernmental organizations (NGOs) have been described as "private, not-for-profit organizations that aim to serve particular societal interests by focusing advocacy and/or operational efforts on social, political and economic goals, including equity, education, health, environmental protection and human rights" (Teegen, Doh, and Vachani 2004, 466). NGOs are an important and influential set of institutional actors within the broad context of the world economy. They have emerged as critical organizations in shaping governmental economic policy and practice, influencing global economic institutions and structures, and affecting corporate and business activities. NGOs have grown in number, power, and influence since the 1980s. They have contributed to a range of international economic debates and issues, and NGO activism has been responsible for major changes in global economic policy, law, and regulation.

According to Lindenberg (1999), fiscal crises, ideological shifts, and privatization have all led to a decline in the scope and capacity of the state. In response, a growing global not-for-profit sector has emerged that, in part, has begun to fill the humanitarian vacuum left by the corporate sector and the nation-state. Teegen, Doh, and Vachani (2004) argue that international NGOs constitute another face of globalization and that large, multinational NGOs have emerged, in part, to serve as counterpoints to concerns about the growing power and influence of multinational enterprises (MNEs) and international organizations such as the International Monetary Fund, the World Bank, and the World Trade Organization.

Civil Society Associations *Civil society*, also referred to as the *third sector* or the *nonprofit sector*, is used broadly to refer to all aspects of society that extend beyond the realm of the public sector and the traditional private sector. Although the term *NGO* is relatively recent, associations among like-minded individuals have been part of ancient and modern history.

When individuals or groups within civil society work together to advance a broad common set of interests, and these interests become a significant force in shaping the direction of society, social movements emerge as the outcomes of this process. Social movements are broad societal initiatives organized around a particular issue, trend, or priority. Modern examples related to economic issues include the environmental movement and the movement to improve working conditions in developing countries. When civil society groups band together to form organized relationships, the emergent entities are often referred to as NGOs.

NGO is a broad term that somewhat loosely refers to all organizations that are neither official parts of government (at any level) nor private, for-profit enterprises. Among NGOs, however, there are many different types, characteristics, and purposes.

Definitions and Classifications Scholars of political economy sometimes separate "club" and "social purpose" NGOs. Club NGOs are membership associations designed primarily to provide a benefit to their members, generally because of pooling interests. Examples of club NGOs are unions, business associations, and church groups.

In most of the contemporary literature the focus is on social purpose NGOs. Social purpose NGOs are engaged in a range of activities that have direct bearing on the world economy, including lobbying global economic institutions such as the WTO to more fully consider the environmental, labor, and developmental impacts of their activities; working with industries and companies to develop codes of conduct governing their activities; providing on-the-ground programs in the areas of health, nutrition, and poverty alleviation; and otherwise supporting institutional developments that promote social welfare.

Teegen, Doh, and Vachani (2004) further differentiate among various functions of NGOs. Advocacy NGOs work on behalf of others who lack the voice or access to promote their interests. They engage in lobbying, serve as representatives and advisory experts to decision makers, conduct research, hold conferences, stage citizen tribunals, monitor and expose actions (and inactions) of others, disseminate information to key constituencies, set and

define agendas, develop and promote codes of conduct, and organize boycotts or investor actions. In these ways, advocacy NGOs give voice and provide access to institutions to promote social gain or mitigate negative spillovers from other economic activity.

Operational (or programmatic or service-oriented) NGOs provide critical goods and services to clients with unmet needs. NGOs have long stepped in to provide critical social safety nets, where politically challenged, indebted, or corrupt states are unable or unwilling to provide for unmet needs, and where global problems defy neat nation-state responsibilities. Examples of such operational activities include relief efforts provided by the Red Cross or Red Crescent, natural resources monitoring by the World Wide Fund for Nature, and the distribution of medicinal drugs by Doctors without Borders. Although some NGOs focus primarily on advocacy or operational service delivery, many others pursue both sets of activities simultaneously, or evolve from one to the other. For example, Oxfam, the global development and poverty relief organization, advocates for changes in public policy that would provide greater support to its efforts while also contributing directly to health, education, and food security in the developing countries in which it operates. Brown and Moore (2001) add another category: capacity-building NGOs, which are large global organizations that use their expertise and financial resources to build the capability of smaller, local NGOs.

The Role of NGOs in the United States and the World Economy NGOs have assumed a significant and influential role in modern societies. The modern era of NGO activism can be traced to 1984, when a range of NGOs, including church and community groups, human rights organizations, and other anti-apartheid activists, built strong networks and pressed American cities and states to divest their public pension funds of companies doing business in South Africa. A U.S. statute banned new U.S. investment in South Africa, export sales to its police and military, and new bank loans, except to support trade. The combination of domestic unrest, international governmental pressures, and capital flight posed a direct,

sustained, and ultimately successful challenge to the white minority rule, resulting in the collapse of apartheid.

NGOs have also pushed to have greater access to trade policy and international governmental agreements and processes, systems that have historically been limited to governments acting as agents of their domestic constituencies. NGOs have expressed a great deal of interest in the trade policy dispute settlement mechanism under the General Agreement on Tariffs and Trade (GATT) and its successor agreement, the World Trade Organization (WTO), as well as in the practices and reform of the International Monetary Fund and the World Bank.

NGOs have also been active in collective efforts to develop, implement, and enforce industrywide standards, codes of conduct, and agreements. Examples of intergovernmental agreements shaped and influenced by NGOs include the WTO and the North American Free Trade Agreement, the United Nations' Global Compact on Business Responsibility, the International Labor Organization's Declaration of Principles Concerning Multinational Enterprises and Social Policy, and the Organisation for Economic Co-operation and Development's guidelines for multinational enterprises (MNEs). Examples of international codes sponsored directly by NGOs include the Social Accountability International SA8000 standard; Rugmark, a standard that certifies rugs and carpets as meeting basic standards for labor and human rights; and the Forest Stewardship Council standard that certifies lumber as consistent with sustainable practices.

According to a 1995 World Bank report, since the mid-1970s, the NGO sector in both developed and developing countries has experienced exponential growth. In terms of international development, more than 15 percent of total overseas development aid is channeled through NGOs. Indeed, a report published by the United Nations and the NGO SustainAbility notes that the global nonprofit sector, with its more than $1 trillion turnover, could rank as the world's eighth largest economy. Teegen, Doh, and Vachani (2004) argue that the emergence of civil society in general, and the activism of civic NGOs in

particular, have broad implications for the role, scope, and definition of corporations in the global economy, and therefore for international management as a research field. Doh and Teegen (2003) point out that the emergence of NGOs, in some cases, has supplanted the role of host governments in the historic business-government bargaining relationship such that NGOs yield significant power over MNEs' right to operate in developing countries.

Another vehicle used by NGOs to advance their agenda is shareholder activism. NGOs may buy shares of corporations and use ownership to promote proxies and other resolutions to effect change. They often use their status to urge institutional shareholders, such as public employee pension and retirement funds, to pursue changes in corporate governance and conduct. They also work with and through socially responsible investment funds, serving as advisors and experts on ethical and social responsibility screens used to determine the composition of such funds and by drawing attention to shortcomings in the mechanisms used by such funds to choose and retain specific stocks within their portfolios.

Challenges to NGOs NGOs face challenges, including the wide diversity of the principals they serve and the constant demands for resources. Variations in national context can also challenge NGOs, as was evidenced by a 2005 Russian law that would appear to limit their ability to mobilize.

In addition, NGOs are facing criticism and pressures over the perception that they are often less accountable for their actions than their government and business counterparts. Specifically, the corporate governance scandals in the United States and around the world have resulted in increasing attention to the role of boards, interlocking board directorates, and overlapping board memberships among corporations and NGOs. The American Enterprise Institute, in cooperation with the Federalist Society for Law and Public Policy Studies, has launched a program initiative called NGO Watch, whose mission is to highlight issues of transparency and accountability in the operations of NGOs and international organizations.

Finally, some argue that, like the MNEs and international organizations they seek to influence, international NGOs are primarily financed and supported by individuals and organizations in wealthy countries and are not as attuned to the interests and needs of the developing world (Lindenberg and Dobel 1999). As a result, many international and local NGOs have developed collaborative relationships that seek to bridge this aspect of the North-South divide.

See also aid, international; anti-globalization; World Bank; World Trade Organization

FURTHER READING

Brown, David L., and Mark H. Moore. 2001. "Accountability, Strategy, and International Nongovernmental Organizations." *Nonprofit and Voluntary Sector Quarterly* 30 (3): 569 87. Discusses and reviews the challenges facing international NGOs, especially as they are called on to fill roles traditionally occupied by governments.

Doh, Jonathan P., and Terrence R. Guay. 2004. "Globalization and Corporate Social Responsibility: How Nongovernmental Organizations Influence Labor and Environmental Codes of Conduct." *Management International Review* 44 (3): 7 30. Discusses the role of NGOs in developing codes of conduct regulating industry.

———. 2006. "Corporate Social Responsibility, Public Policy, and NGO Activism in Europe and the United States: An Institutional Stakeholder Perspective." *Journal of Management Studies* 43: 47 73. Compares and contrasts NGO activism in Europe and North America.

Doh, Jonathan P., and Hildy Teegen, eds. 2003. *Globalization and NGOs: Transforming Business, Government, and Society*. Westport, CT: Praeger. Comprehensive edited volume with contributions from scholars and practitioners.

Guay, Terrence R., Jonathan P. Doh, and G. Sinclair. 2004. "Nongovernmental Organizations, Shareholder Activism, and Socially Responsible Investments: Ethical, Strategic, and Governance Implications." *Journal of Business Ethics* 52: 125 39. Discusses the role of NGOs in corporate social responsibility, especially socially responsible investment.

Lindenberg, Mark. 1999. "Declining State Capacity, Vol untarism, and the Globalization of the Not for Profit Sector." *Nonprofit and Voluntary Sector Quarterly* 28: 147 67.

Lindenberg, Mark, and Patrick J. Dobel. 1999. "The Challenges of Globalization for Northern International Relief and Development NGOs." *Nonprofit and Vol untary Sector Quarterly* 28: 4 24.

Teegen, Hildy, Jonathan P. Doh, and Sushil Vachani. 2004. "The Importance of Nongovernmental Organizations (NGOs) in Global Governance and Value Creation: An International Business Research Agenda." *Journal of International Business Studies* 35: 463 83. Provides a broad review of the role of NGOs in global business affairs.

United Nations. 2006. "UNIC and Nongovernmental Organizations (NGOs)." Downloadable from http://www.un.org.pk/unic/ngo.htm. Reviews the ways in which the UN interacts with NGOs.

World Bank. 1995. *Working with NGOs*. Washington, DC: The World Bank. Provides a guide to World Bank di visions and affiliates on how to work with NGOs.

JONATHAN P. DOH

■ nontariff measures

Nontariff measures (NTMs) are policies, rules, regulations, and practices *other than tariffs* that distort international trade flows. Tariffs are taxes imposed on imported goods as they cross borders and clear customs. Any other measures that, intentionally or unintentionally, distort international trade flows are considered nontariff measures. Most NTMs are nontariff trade *barriers* (NTBs) that reduce trade flows. Import quotas, countervailing or anti-dumping duties, and import surcharges are NTBs specifically designed to restrict international trade. Other government regulations, such as sanitary requirements or product safety standards, are directed at protecting public health and safety, but become NTBs if they also restrict imports. Not all NTMs restrict trade. *Export subsidies,* for example, distort trade flows, but increase rather than decrease international trade.

The most important of the vast array of potential NTMs can be categorized as (1) taxlike measures, (2) cost-increasing measures, (3) quantitative trade measures, or (4) government procurement policies.

Taxlike Nontariff Trade Measures Tariffs usually apply to imports from all sources and usually remain in effect at the same rates for many years. Some taxlike NTMs apply only in special circumstances to imports from selected countries and/or are only temporary.

Anti-dumping duties are additional import taxes imposed in response to the "dumping" of products by foreign firms. Dumping occurs when an exporter sells abroad at a price below the price charged for the same product in the home market, or sells abroad at a price below its average cost of production. Importing countries levy anti-dumping duties only on imports of the product from the country of the firm found to be dumping. The anti-dumping duty is set at a rate equal to the "dumping margin," which is the difference between the price the firm charges in the home market and in the importing country, or the difference between the exporter's average cost of production and the price charged in the importing country. Anti-dumping duties are imposed in addition to the standard tariff on the product.

Countervailing duties are extra import taxes imposed on goods receiving subsidies from the exporting country's government. The size of the countervailing duty normally equals the size of the subsidy. Countervailing duties also are imposed in addition to the regular tariff.

Temporary import surcharges are extra import taxes imposed in "emergency" circumstances, often in response to a sudden increase in a country's balance of payments deficit. For example, the United States imposed a temporary 10 percent import surcharge during the second half of 1971 following several years of growing balance of payments deficits.

Variable levies are taxes on imports that vary in size depending on the price of the imported product in a particular shipment. For example, the European Community's system of variable levies on agricultural imports imposed an import tax just large enough to eliminate any difference between the internal

support price and the price of the imported shipment. The variable levies ensured that agricultural goods could not enter the EC at prices lower than domestic support prices.

Cost-Increasing Nontariff Trade Barriers Some NTBs do not directly tax goods, but discourage imports by increasing either the exporter's cost of production or the cost of clearing customs and complying with import regulations.

Prior import deposits are requirements that importers submit deposits to the importing country's Central Bank equal to a percentage of the value of the shipment. These deposits usually are denominated in the importing country's currency, pay no interest, and are held for about 6 months. For example, Uruguay at one point required importers of certain products to submit prior import deposits of up to 1200 percent of the value of the shipment that were held for 6 months at no interest when the inflation rate was about 50 percent per year. Prior import deposits can create delays and impose extra costs on importers.

Product standards, technical regulations, labeling requirements, and sanitary and phytosanitary requirements make imported goods more expensive by increasing an exporter's cost of production to comply with the standard, or by imposing costs on importers due to shipment delays or inspection or certification requirements. For products subject to economies of scale in production, altering a product to comply with widely differing product standards across countries can preclude cost-reducing long production runs. Importers may also have to turn over a significant portion of a shipment for testing. Technical standards such as the European Union's (EU's) import ban on beef that has been fed growth hormones and the French ban on imports of products containing asbestos completely prohibit trade in those products.

Customs procedures that are costly, cause lengthy delays, or create uncertainty for importers discourage trade. International traders complain that inspection and customs clearance procedures in many countries are time consuming and costly. Importers may hesitate to import if they are uncertain as to how goods will be classified and thus what tariff rate will apply.

Reference prices or minimum import prices are official prices used to value import shipments to calculate the amount of duty owed. The usual basis for valuation is the amount the importer pays the exporter, including insurance and freight charges. Reference prices increase the amount of duty paid by an importer by using arbitrary (and higher) prices to value shipments. The United States for many years valued imports of benzenoid chemicals, rubber-soled footwear, woolen knit gloves, and clams at the (higher) internal "American selling price" instead of the actual cost of the shipments.

Quantitative Measures In contrast to tariffs, taxlike NTMs, or other measures that discourage imports by making them more expensive, quantitative trade restrictions directly or indirectly restrict the *quantity* of a product that can be imported or exported.

An *import quota* sets a maximum permissible volume or value of a product that can enter a country during a particular time period. The import quota limit is normally declared before the beginning of the time period. The overall quota limit is often subdivided by country of origin. Import quotas can be administered on a "first-come, first-served basis" by allowing imports until the quota ceiling is reached and then closing the border to further imports. However, governments normally distribute licenses or permits to import among importing firms. The distribution is most often determined by the amounts imported in a base year. *Export quotas* are maximum limits on the amount of a product that can be exported during a particular period of time.

Embargos are total prohibitions on trade of a particular good or trade with a particular trading partner. Most countries prohibit trade in products originating from endangered species or trade in weapons or radioactive materials. Embargos on trade with particular trading partners are usually for political reasons, such as the U.S. embargo on trade with Cuba or the United Nations embargo on trade with South Africa during the period of apartheid.

A *voluntary export restraint* (VER) is an export quota on shipments of a product to a particular importing country that is "voluntarily" applied by the exporting country, usually as the result of negotiations with the importing country. Negotiating VERs with exporting countries was a way governments could protect troubled domestic industries when they could not unilaterally increase trade barriers without violating existing international trade agreements. Exporting country governments often agreed to restrict exports to avoid even more severe unilateral import restrictions by the importing country. Major VERs included European and Japanese limits on steel exports to the United States, Japanese VERs on automobile exports to the United States and European countries, and the complicated web of VERs on textile and apparel products under the Multifiber Arrangement (MFA).

Import licensing requirements are regulations that prohibit importation of designated products without a license issued by a government agency. The agency in charge may base the number of import licenses approved on foreign exchange availability and the state of the overall balance of payments. But the authorities granting import licenses often have a high degree of discretion as to which products can be imported and which importers receive licenses. *Export licensing requirements* are the equivalent restrictions on the export side.

Foreign exchange controls are regulations requiring that exporters turn over any foreign exchange (foreign currency) earnings to the Central Bank in exchange for domestic currency. The Central Bank then allocates the foreign exchange among importers. Importers cannot import without access to foreign exchange to pay for the goods, so foreign exchange allocation decisions limit import and also determine which goods can be imported.

A *tariff quota*, or *tariff rate quota*, is a hybrid form of trade restriction that allows imports at a "within-quota" tariff rate up to a maximum quantity (the "quota" part of the tariff quota). Additional imports "over-quota" are allowed but are subject to a higher tariff rate. If imports fall short of the quota limit, then the impact of a tariff quota is similar to that of a tariff at the "within-quota" rate. If imports enter "over-quota," then the impact of the tariff quota is similar to that of a tariff at the "over-quota" rate. But if the over-quota tariff rate is so high that no imports enter in excess of the quota limit, then the tariff quota acts like an import quota. The United States, Canada, and the EU apply tariff quotas to some agricultural imports.

Local (or "domestic") content requirements indirectly limit imports of components by mandating that goods sold on the domestic market must contain a minimum percentage of local value added or a minimum percentage of components sourced from local producers. Mexico, Brazil, Argentina, Uruguay, Peru, and the Philippines have at one time required minimum local content in motor vehicle or household appliance industries. *Mixing regulations* or *quantity-linking schemes* require importers to purchase a minimum amount of domestic product before they are permitted to import that product from abroad. An example would be the requirement in Mali that sugar importers purchase one ton of sugar from the domestic sugar refinery for each ton of sugar that they import. Unless the country in question is granted a special exemption, these measures are now prohibited for World Trade Organization members under the Agreement on Trade-Related Investment Measures.

Import/export balancing requirements, sometimes called *compensatory export requirements*, force firms to export to earn foreign exchange equal to a specified percentage of the value of the goods they import. These regulations often apply to motor vehicle assembly. Firms must export motor vehicles or parts to earn foreign exchange to pay for a given percentage of their imported components.

Government Procurement Practices In most countries, regardless of the stage of development, government is the single largest purchaser of goods and services. Purchases of goods and services by governments represent a large potential market for foreign suppliers. Government procurement practices include the method of soliciting bids,

requirements placed on bidders, and the method of selecting winning bids and awarding contracts. Discrimination against foreign suppliers could take the form of allocating contracts exclusively to domestic firms, not allowing foreign bidders to participate in procurement solicitations, not providing adequate information to potential foreign bidders, or establishing preference margins for domestic firms. In the United States, for example, the Buy American Act of 1933 required government agencies to purchase only products produced domestically, unless the cost would be "unreasonable." The interpretation of "unreasonable" gave domestic firms a preference margin as high as 50 percent.

Comparison of Tariffs and NTMs Tariffs increase the prices of imported goods and shift demand toward domestically produced substitutes. Consumers bear the burden of higher prices of both imported and domestic goods. Domestic producers benefit because they sell a greater quantity at a higher price. Tariffs also provide revenue to the government. Tariffs generally result in net losses to the importing country overall because the loss to consumers from higher prices exceeds the gains to domestic producers and the government.

NTBs also increase the prices of imported goods and domestic substitutes. The taxlike NTMs provide the government with revenue, but cost-increasing NTMs raise prices to consumers without transferring revenue to the government. Quotas provide windfall profits to whoever is able to trade the good under quota. Depending on how the quotas are administered, the windfall gains may be captured by exporters, which would increase the net loss to the importing country from restricted trade.

In general, NTMs are less transparent, more discriminatory, and more discretionary than tariffs. Tariffs are said to be *transparent* because the amount by which a tariff increases the price of an imported good is clear. A 20 percent tariff increases the cost of a shipment to the importer by 20 percent. The extra cost imposed by customs procedures, inspections, and compliance with standards is far more difficult to quantify. Quantitative trade restrictions also are less transparent than tariffs because the price-increasing

effect of an announced quota ceiling expressed in tons or square yards is not immediately apparent. With tariffs it is also clear that the government receives the tariff revenue. Quantitative restrictions generally provide no government revenue, but create profits for the limited number of traders who are able to import within the quota.

The welfare impact of NTMs such as product safety standards, sanitary requirements, and labeling requirements is particularly difficult to assess. These measures generate social benefits (such as improved health, safety, or improved information for consumers) that can outweigh the costs. To the extent that these measures reduce consumer uncertainty about the quality of imported goods, they may increase, rather than decrease, international trade.

NTMs also provide more scope for *discrimination* against particular exporting countries. Tariffs normally apply at the same rate to products from most exporting countries. Import quotas, however, are often subdivided among exporting countries. VERs are more often negotiated with countries that are viewed as source of increased imports, or countries over which the importing country has more political leverage. Inspections to assure compliance with sanitary or product standards can be applied more frequently or more stringently to products from particular sources.

NTMs are often more *discretionary* than tariffs. Tariffs apply to all shipments at a preestablished rate. Import licensing, foreign exchange control systems, or import quotas give government authorities much more discretion to control what can be imported, when it can be imported, from which exporting countries, and by which importers.

NTMs in International Trade Agreements and Negotiations Two important fundamental principles of the General Agreement on Tariffs and Trade (GATT) and its successor, the World Trade Organization (WTO), are the use of tariffs as the primary form of trade barrier, and nondiscriminatory treatment of imports from all parties to the agreement. Successive rounds of trade negotiations between 1947 and 1994 under the auspices of the GATT made remarkable progress in lowering tariff rates on

trade in manufactured goods. Following the principle of nondiscrimination, the lowered tariff rates applied to imports from all exporting countries that were parties to the GATT.

As tariff rates fell, the trade-restricting effect of NTBs became increasingly apparent. Lower tariff rates made preexisting NTMs more evident. But lower tariffs also prompted import-competing industries to seek additional protection through anti-dumping and countervailing duties, VERs, and technical barriers to trade. The perceived proliferation of NTMs prompted governments to try to negotiate limitations on their spread and rules on their application.

The first negotiated agreement on NTMs was the Anti-dumping Code that emerged from the Kennedy Round (1963 67). Subsequent rounds of trade negotiations put increased emphasis on negotiations to reduce the trade-distorting effect of NTMs. The Tokyo Round (1973 79) resulted in the negotiation of ancillary codes on import licensing procedures, technical barriers to trade, customs valuation, subsidies and countervailing duties, and government procurement. The Uruguay Round (1986 94) further strengthened the rules governing the use of NTMs by specifically prohibiting VERs, replacing agricultural NTMs (such as U.S. import quotas and the EU variable levies) with tariffs or tariff rate quotas, and incorporating agreements on anti-dumping duties, subsidies and countervailing duties, and sanitary and phytosanitary measures into the main WTO agreements. An agreement was also reached in the Uruguay Round to totally phase out the Multifiber Arrangement, arguably the worst of all nontariff measures, by 2005. Despite these changes, however, NTMs still play an important role in the world trading system, and understanding their impacts is an ongoing area of study for trade policy analysts and researchers.

See also anti-dumping; countervailing duties; government procurement; political economy of trade policy; quotas; tariff rate quotas; tariffs; technical barriers to trade; trade facilitation; trade-related investment measures (TRIMs)

FURTHER READING

Baldwin, Robert E. 1970. *Nontariff Distortions of International Trade*. Washington, DC: The Brookings Institution. An early and thorough survey of the different types of nontariff trade barriers.

Corden, W. M. 1971. *The Theory of Protection*. Oxford: Oxford University Press. Contains analyses of mixing regulations (quantity linking schemes), content protection, and an extensive analysis of quotas.

Deardorff, Alan V., and Robert Stern. 1998. *Measurement of Nontariff Barriers*. Ann Arbor: University of Michigan Press. Describes the difficulties involved in measuring the impact of various types of NTMs and cites previous studies. Also includes an appendix with a somewhat different method of categorizing NTMs.

Finger, J. Michael, ed. 1993. *Antidumping: How It Works and Who Gets Hurt*. Ann Arbor: University of Michigan Press. A collection of essays on the origins of anti dumping actions and case studies of anti dumping actions. Also includes enforcement experience in Australia, Canada, the United States, and the European Community.

Hillman, Jimmye S. 1978. *Nontariff Agricultural Trade Barriers*. Lincoln: University of Nebraska Press. A early survey of NTBs in agriculture with case studies.

Laird, Sam, and Alexander Yeats. 1990. *Quantitative Methods for Trade Barrier Analysis*. New York: New York University Press. Analyzes the economic impact of selected NTMs, discusses the methods of estimating their impact, summarizes previous studies by country and product, and includes a glossary of NTMs.

Maskus, Keith E., and John S. Wilson, eds. 2001. *Quantifying the Impact of Technical Barriers to Trade: Can It Be Done?* Ann Arbor: University of Michigan Press. Includes chapters that model the impact of technical barriers to trade and review attempts to quantify their impact.

Vousden, Neil. 1990. *The Economics of Trade Protection*. Cambridge: Cambridge University Press. Compares tariffs and quotas and contains analyses of VERs, local content schemes, variable levies, and export subsidies, with descriptions and explanations.

WENDY TAKACS

■ nontraded goods

Even in today's increasingly integrated world a host of goods and services — health care and housing, for example — are not traded across national borders, either because they simply cannot be traded or because barriers to trade are too high. Sometimes political barriers also may prevent trading of some goods, such as strategic military equipment. These commodities, traded exclusively at home (domestically), are broadly labeled as nontraded goods.

Nontraded goods must have their markets cleared locally, and in this respect, differ fundamentally from traded goods for which local excess demand or supplies can be accommodated in the world markets. Although recent trends in globalization, characterized by sharp declines in transportation and communication costs, tend to convert nontraded goods into traded goods, nontraded goods and services continue to represent the large majority of world consumption.

Goods that are traded internationally can be transported or provided in another country at some cost. In principle, therefore, the providers of such goods in different countries compete with one another. In contrast, nontraded goods have to be provided locally and do not involve competition among international providers. The classic textbook example of a nontraded good is a haircut (albeit a service). Though substantial differences may exist in the prices of haircuts across countries, trade does not develop by shipping haircuts from low-price to high-price countries. The cost of transporting either the barber or the customer to take advantage of the low price would far outweigh the price differential. As such, a barber in the United States need not be excited to learn that an appreciation of the pound has made haircuts in the United Kingdom relatively costlier.

The importance of nontraded goods in a trading world was discovered by Ohlin (1929) in his criticism of Keynes's (1929) views on the German transfer problem and has subsequently been illuminated by many prominent economists. Keynes and Ohlin agreed that a transfer would inevitably result in a reallocation of resources between the traded and nontraded sectors. Keynes had argued, however, that the presence of a nontraded good would shift resources away from the nontraded sector into the exporting sector to exacerbate the donor's terms of trade deterioration as a result of a transfer. Ohlin, in contrast, saw no reason for a movement in the terms of trade and expected only opposing changes in the prices of traded goods relative to those of nontraded goods for the donor and recipient.

Prices of nontraded goods are determined exclusively by local conditions that affect demand and supply, whereas prices of traded goods are determined by conditions around the globe. As such, the distinction between traded and nontraded goods becomes even more important for relatively small trading countries: a small country, through policy changes, can affect prices of nontraded goods but it cannot affect prices of traded goods.

Theories of Trade and Nontraded Goods Are the results of the models of international trade sensitive to the presence of nontraded goods? Jones (1974) presented an elegant model, in which he assumed away the possibility of import-competing production and local consumption of exports, providing a clear view of the way in which the market for nontraded goods is linked to that for traded goods. Komiya (1967) was the first to generalize the standard Heckscher-Ohlin (H-O) model for international trade to incorporate a third, nontraded good. The presence of a nontraded good removes the production indeterminacy of a two-factor, three-commodity (two traded and one nontraded) H-O model. The factor price equalization (FPE) theorem and Rybczynski theorem on the effect of a change in factor endowments on the output levels hold in the new three-commodity model. A country's trade bias (pro or anti) depends on the elasticity (or inelasticity) of the demand for the nontraded good as well as the relative factor intensities of all three commodities. Technological progress in the traded sector does not necessarily result in an increase in its output. Technological progress in the nontraded sector, however, raises the output of the nontraded goods sector if inferior goods are ruled out.

Ethier (1972) analyzed the effects of integrating nontraded goods on each of the four basic results of

the H-O model: the FPE theorem, the Rybczynski theorem, the Stolper-Samuelson theorem, and the proposition that a change in the terms of trade will cause an increase in the output of the good whose relative price has risen and a decrease in the output of the other when there are fixed-factor endowments. It can also be shown that the Rybczynski theorem for net output movements stays unaffected by interindustry flows in the two-factor, two-commodity Heckscher-Ohlin model for gross and net output movements in the presence of traded and nontraded intermediate inputs. If the nontraded commodity is a pure intermediate commodity, the Rybczynski theorem still holds with some modifications for gross and net output movements.

In a two-factor, three-commodity model where the third commodity is a nontraded good, the Rybczynski theorem holds for output movements if there is no inferiority in consumption. This is true with some modifications for gross and net output movements when interindustry flows exist. A sufficient condition for this generalized version of the Rybczynski theorem, applicable to a two-factor, three-good economy where one of the goods is nontraded and indecomposable interindustry flows exist, is that the net output change of the nontraded good is bounded by the factor changes.

Following Rivera-Batiz (1982), Komiya's results on the effect of technological changes and changes in factor endowments in the presence of nontraded goods do hold, in a framework where the number of factors and the number of goods are equal, if and only if the sum of the own price elasticities of demand for traded and nontraded goods is equal to unity. Otherwise, a Hicks-neutral technological change in the nontraded goods sector will reduce the price of nontraded goods proportionally less (more) in absolute value than the technological change when the aggregate elasticity is greater (smaller) than unity. Factor prices measured in terms of the traded goods will in general not be fixed, and factor prices measured in terms of the nontraded goods will not change proportionally to the same extent as the change in the price of the nontraded goods, as long as the aggregate elasticity is different from one. The

exact effect on the factor prices will depend in the magnitude of the aggregate elasticity and on the factor-intensity of the nontraded goods' production relative to the traded goods' production. Also, factor prices and the price of nontraded goods are affected by changes in factor endowments in this framework. If the endowment of any factor increases, the price of nontraded goods will decrease if the nontraded goods production is relatively intensive in the expanding factor. Income distribution will turn against the expanding factor. In the context of the specific-factors model, in which an extreme asymmetry in factor mobility is assumed, Clague (1985) showed that increases in the endowment of a specific factor increases the real reward of the mobile factor when nontraded goods exist.

Adding nontraded goods to a conventional factor proportions model, according to Deardorff and Courant (1990), will reduce the likelihood of FPE. Specifically, in the presence of nontraded goods, the reduction in the likelihood of FPE is equal to the fraction of income spent on nontraded goods. For given prices, the set of factor endowment combinations that are consistent with producing the same set of goods and having the same factor prices will shrink, reducing the chance that a country's factor endowments lie in the set, and therefore the likelihood of FPE is reduced. A generalization of the Stolper-Samuelson theorem to include the role of the nontraded goods sector in explaining the effects of price changes in traded goods on relative and absolute wages, provided by Beladi and Batra (2004), reveals that if the import and nontraded sectors are reduced, then freer trade can harm skilled and unskilled labor while improving capital.

Some of the standard conclusions about international trade are disrupted when market imperfection is present in the nontraded goods sector. The fact that some commodities are not traded internationally may itself suggest the presence of product market imperfections. Prohibitive tariffs, for instance, are a source of nontraded goods that may confer market power. In the presence of a monopolized nontraded good, the Rybczynski theorem remains intact as long as the price elasticity of demand for nontraded goods

depends only on relative prices. Following Cassing (1977), however, the Stolper-Samuelson theorem is modified in an important way depending on the degree of monopoly and the specific taste patterns of the factors. An increase in the relative price of a traded good will raise the nominal reward for the factor used intensively in that sector but will lower the price elasticity of demand in the nontraded sector which, in turn, will raise the price of the nontraded good relative to the gaining factor whose nominal reward improved. The real reward of the factor, used intensively in the traded sector experiencing a price increase, can fall for sufficiently high propensity of nontraded goods consumption by that factor. Within the same framework, Hazari and Kaur (1995) show that when a nontraded good is produced by a monopoly, a tourism boom can lead to a decrease in the domestic residents' welfare because of drastic movements in the terms of trade.

Policy Relevance of Nontraded Goods The existence of nontraded goods has nontrivial implications for a country's domestic or international policies. An expansionary fiscal policy in nontraded goods and an expansionary monetary policy in international bonds can have offsetting effects on a country's reserves. When the production possibility set is strictly convex (i.e., there are no economies of scale), if the highest tariff rate is reduced to the level of the next highest, welfare improves under certain conditions: specifically, no inferior goods can exist in the economy, the commodity with the highest tariff rate must be substitutable, and the nontraded goods must be substitutable. When multiple commodities share the highest tariff rate, and the highest tariff rate is reduced to the level of the next highest rate, the level of utility increases under substitutability. When all tariff rates move proportionally to a given rate (uniform tariff change), utility increases. Tariffs and production subsidies can reduce the price of nontraded goods. A consumption subsidy may be the optimal policy decision in a model with nontraded goods and production externalities.

When there exists a high intertemporal elasticity of substitution in consumption, contrary to the case with traded goods only, a temporary fiscal expansion (in the form of government spending) on nontraded goods causes real interest rates to move in different directions in each country. More specifically, a fiscal expansion of nontraded goods with a low elasticity of substitution will decrease the domestic interest rate relative to world real interest rates, deteriorating the expansionary country's current account. With a high substitution elasticity, however, the expansion increases domestic real interest rates relative to foreign real interest rates while improving the domestic country's current account. In the presence of a generalized cash-in-advance constraint, when a monetary distortion of the nontraded good is present, a tariff may not be dominated by a consumption tax.

In a two-sector economy, if one of the sectors is characterized by monopolistic competition, a tariff leads to an improvement in welfare if the differentiated good is nontraded but leads to a deterioration of welfare if the homogeneous good is nontraded. Decreases in import prices or tariffs will increase the price of nontraded goods when cross-substitution effects are dominated by the income effect. In a setting of differentiated traded goods, welfare decreases as a result of the tariff causing a demand shift away from the differentiated goods market toward the market for the nontraded homogeneous good. Devereux (1988) offers an insightful discussion on these issues.

Welfare Analysis and Nontraded Goods The relevance of nontraded goods for welfare analyses, including but not limited to the "transfer paradox" (i.e., the possibility that a transfer may reduce the recipient's welfare even when markets function perfectly), has added interest in the topic. Early attempts at identifying a distinct role of nontraded goods in welfare convincingly argued that the effects of neutral technical progress on output and welfare are sensitive to the inclusion of nontraded goods. The welfare of the host country can be shown to unambiguously improve with foreign investment in nontraded goods. The gains-from-trade proposition applies to a monopolistic competition framework that extends to include nontraded goods when differentiated nontraded goods are present as well as traded goods.

Technical progress in manufacturing increases the price of nontraded goods and improves national welfare, and deterioration in the terms of trade increases the price of the nontraded goods and improves national welfare, if the gain of the induced employment effects exceeds the direct loss of the terms-of-trade decline. Those who own traded goods suffer welfare deterioration while owners of nontraded goods experience welfare improvement following a temporary devaluation.

It is important to realize that international trade grows not only as countries trade (on the intensive margin) more of the goods they had already been trading but also when countries begin trading goods (on the extensive margin) they had previously not been trading (nontraded goods). As such, it is hardly surprising that nontraded goods continue to draw our attention even in today's increasingly globalized world.

See also gains from trade; Heckscher-Ohlin model; real exchange rate; tariffs; terms of trade

FURTHER READING

Beladi, Hamid, and Ravi Batra. 2004. "Traded and Non Traded Goods and Real Wages." *Review of Development Economics* 8 (1): 1 14. Generalizes the Stolper Sa muelson theorem to include the role of nontraded goods.

Cassing, Jim. 1977. "International Trade in the Presence of Pure Monopoly in the Non Traded Goods Sector." *Economic Journal* 87 (347): 523 32. Provides one of the most insightful analyses of the role of the nontraded sector with market power.

Clague, Christopher K. 1985. "A Model of Real National Price Levels." *Southern Economic Journal* 51 (4): 998 1017. Analyses the real return to the mobile factor in the presence of nontraded goods.

Deardorff, Alan V., and Paul N. Courant. 1990. "On the Likelihood of Factor Price Equalization with Non Traded Goods." *International Economic Review* 31 (3): 589 96. Provides implications of nontraded goods for factor price equalization.

Devereux, Michael. 1988. "Non traded Goods and the International Transmission of Fiscal Policy." *Canadian Journal of Economics* 21 (2): 265 78. Contains an in sightful analysis of differentiated good market and nontraded goods.

Ethier, Wilfred. 1972. "Nontraded Goods and the Heck scher Ohlin Model." *International Economic Review* 13 (1): 132 47. Contains some of the most effective ar guments raising concerns regarding the implications of a model of nontraded good with an equal number of goods and factors.

Hazari, Bharat R., and Charanjit Kaur. 1995. "Tourism and Welfare in the Presence of Pure Monopoly in the Nontraded Goods Sector." *International Review of Economics and Finance* 4 (1995): 171 77. Explores tourism booms under a monopolized nontraded sector.

Jones, Ronald W. 1974. "Trade with Non traded Goods: The Anatomy of Inter connected Markets." *Economica* 41 (162): 121 38. Offers one of the most elegant for mulations of the link between traded and nontraded goods.

Keynes, John M. 1929. "The German Transfer Problem." *Economic Journal* 39 (153): 1 7. Contains one of the earliest assessments of the importance of nontraded goods.

Komiya, Ryutaro. 1967. "Non Traded Goods and the Pure Theory of International Trade." *International Economic Review* 8 (2): 132 52. Provides one of the most general treatments of a trade model with nontraded goods.

Ohlin, Bertil. 1929. "The Reparation Problem: A Discus sion. I. Transfer Difficulties, Real and Imagined." *Economic Journal* 39 (154): 172 78. Shaped one the most compelling criticisms of Keynes' assessment of the sig nificance of nontraded goods in a trading world.

Rivera Batiz, Francisco L. 1982. "Non traded Goods and the Pure Theory of International Trade with Equal Numbers of Goods and Factors." *International Economic Review* 23 (2): 401 9. Contains an analysis of the effects of changes in factor endowments on the factor prices in the presence of nontraded goods.

HAMID BELADI AND AVIK CHAKRABARTI

■ North American Free Trade Agreement (NAFTA)

The United States, Canada, and Mexico are signatories to the North American Free Trade Agreement (NAFTA). In 2006, the pact covered a combined

economy of more than $14 trillion and a population of about 435 million people. In large measure, NAFTA represents the agglomeration of Mexico into the existing Canada United States Free Trade Agreement (FTA) that was signed in January 1988 and became effective a year later. Mexico initially sought a bilateral FTA with the United States; the two countries announced their intent to pursue an FTA in June 1990, which evolved into NAFTA when Canada joined the process several months later. Negotiations began in May 1991 and concluded in August 1992; U.S. president George H. W. Bush, Mexican president Carlos Salinas, and Canadian prime minister Brian Mulroney signed the agreement in December 1992.

Prior to its ratification, however, the incoming U.S. president, William J. Clinton, demanded additional commitments regarding labor, the environment, and import safeguards. Subsequent negotiations during the first eight months of 1993 produced three side agreements to NAFTA: the North American Agreement on Labor Cooperation (NAALC); the North American Agreement on Environmental Cooperation (NAAEC); and the Understanding between the Parties to the NAFTA Concerning Chapter 8 Emergency Action (that is, against import surges). In addition, the United States and Mexico signed the Border Environmental Cooperation Agreement, which, inter alia, created the North American Development Bank with the aim of funding environmental projects on both sides of the border. Together, these "side accords" supplemented NAFTA provisions and helped Clinton secure congressional approval of the pact in November 1993 by a vote of 234 to 200 in the House and 61 to 38 in the Senate. NAFTA entered into force on January 1, 1994.

Why NAFTA? NAFTA arose because Mexico needed and asked for free trade with the United States. Given the legacy of the debt crisis of 1982, low domestic savings, and an overvalued peso, Mexico needed to import technology and capital to propel economic growth. By ensuring open access to the U.S. market and making it harder to reverse Mexican reforms implemented since the *apertura* (opening) of the 1980s, Mexico expected NAFTA to complement continuing domestic reforms and create new trade and investment opportunities within the Mexican economy. In fact, the mere announcement of NAFTA talks elicited significant new commitments of foreign direct investment in Mexico in anticipation of the new trade regime with the United States.

For the United States, NAFTA served both economic and political objectives. NAFTA created new opportunities to increase trade and investment with its southern neighbor; U.S. officials also touted though greatly exaggeratedly the employment implications of the trade pact. At the same time, they saw the initiative as a way to support the growth of political pluralism and democracy in Mexico, and as part of the long-term response to chronic immigration problems.

Overall, NAFTA has had only a glancing effect on these nontrade issues. NAFTA's first decade coincided with profound changes in Mexico's political regime and the end of 70 years of one-party rule. Mexico's political reform has been a uniquely Mexican achievement, however. NAFTA deserves little credit, except to the extent that increased U.S. interest in Mexico made the 2000 election in Mexico harder to rig. On immigration, NAFTA has not met expectations. Migration from Mexico, both legal and illegal, increased over NAFTA's first decade. Demographers initially forecast increased flows of Mexican migration to the United States, followed by declines in the medium and long run as Mexican economic prospects brightened. But these estimates had to be recalibrated in light of the peso crisis, enduring poverty in states south of Mexico City, persistent rural-urban migration, and the impact of Chinese competition on Mexican industry.

What Is NAFTA? NAFTA was the first free trade agreement involving developed and developing countries that established a comprehensive set of obligations on trade and investment in goods and services. Its detailed chapters cover subjects such as customs procedures, technical standards, sanitary and phytosanitary measures, government procurement, intellectual property rights, investment and

competition policy, and sectors such as agriculture, energy, and cross-border services trade in telecommunications and financial services.

With a few notable exceptions, the pact removes barriers to trade and investment in goods and services among the three economies. It does not create a customs union; each member maintains an independent trade regime governing its transactions with nonmember countries. On agriculture, a sector often excluded from disciplines in trade accords, NAFTA comprises three bilateral pacts (U.S.-Mexico, U.S.-Canada, and Canada-Mexico). The U.S.-Mexico accord comes closer to free trade in farm goods than any other extant trade agreement and provides more pervasive liberalization of trade barriers than the U.S.-Canada pact, which contains important exceptions to the free trade regime originally included in the Canada-U.S. FTA that were carried over into NAFTA.

In addition, NAFTA incorporates six dispute settlement processes to manage and expedite the resolution of disputes among the three countries. The six processes cover chapter 11 (investment), chapter 14 (financial services), chapter 19 (anti-dumping and countervailing duties), chapter 20 (functioning of the agreement), the NAALC (labor), and the NAAEC (environment). Chapter 19 cases have been by far the most numerous; chapter 11 provisions involving investor-state disputes have been the most controversial, contrary to initial expectations. When these investor rights were first conferred, the chapter 11 provisions were hailed as a better forum than national courts for resolving investment disputes. In practice, however, the rules (e.g., the ban on indirect expropriation under article 1110 and the minimum standards under article 1105) have fostered litigation by business firms against a broader range of government activity than originally envisaged. Environmental groups, in particular, have raised concerns about the impact of chapter 11 tribunals on state and local environmental standards.

By contrast, NAFTA's institutional structure is skeletal. The signatories did not want to create a new bureaucracy; the governance structure of the pact is minimalist and its budget insufficient to achieve even its modest mandate. The pact is governed by a Free Trade Commission composed of the three trade ministers that oversee the operation of the agreement. In addition, there are more than 30 working groups that provide advice to the commission on specific issues (e.g., rules of origin), a small NAFTA Secretariat to administer the dispute settlement procedures and other work of the commission and working groups, and separate advisory bodies for the labor and environment accords.

Although the pact is comprehensive compared with other trade agreements, it has important limitations. First, NAFTA omits important rights and obligations on subsidies, energy trade, and investment, and the use of anti-dumping and countervailing duties. The pact also maintains restrictions on U.S.-Canadian agricultural trade excluded from their bilateral FTA. In addition, its side pacts on labor and the environment are primarily consultative arrangements and not backed by meaningful arbitration mechanisms or financial resources.

Second, NAFTA was not designed to cure the manifold ills of North American societies, including high levels of illegal immigration, slow progress on environmental problems, growing income disparities (particularly within Mexico), weak growth in real wages, and trafficking in illegal drugs. Some of these problems are correlates of economic integration and higher incomes; all of them require policy responses that range well beyond the competence of trade officials.

The NAFTA Experience to Date In commercial terms, NAFTA has met the stated objectives of its signatories but failed to justify either the inflated promises of its political supporters or the harsh defamations of its critics. Over its first decade, intra-NAFTA trade soared, building on the extensive network of cross-border investments already linking, in particular, the auto and electronics industries of North America. Merchandise trade among the three countries rose almost threefold and was valued at more than $800 billion in 2006; exports to and imports from Canada and Mexico accounted for 30 percent of total U.S. trade. Direct investment by the

United States in Mexico grew robustly but still represented only a small fraction of total investment in American plant and equipment by U.S. firms. The stock of U.S. holdings in Mexico was valued at more than $71 billion at year-end 2005, up from $17 billion in 1994. Canada-Mexico trade and investment also grew rapidly from a low base.

Net employment rose sharply in all three countries during NAFTA's first decade, aided by the positive add-on effects of strong U.S. economic growth. Millions of jobs were created in each country and a sizable but smaller number of jobs were lost each year. On balance, employment rose by 15 million in the United States, 3 million in Canada, and 8 million in Mexico. To be sure, employment in Mexican maquiladoras fell sharply from peak levels in 2000, but those losses resulted primarily from the U.S. economic downturn in 2001 2, changes in Mexican tax policy, and the impact of a strong peso on industrial competitiveness. In the United States, NAFTA-related job losses represented a small fraction of displaced workers whether one counts the official tally of the NAFTA Trade Adjustment Assistance program (about 60,000 per year) or the higher numbers cited by NAFTA critics.

Similarly, NAFTA's impact on wages got swamped by other factors in the massive U.S. economy. Claims that NAFTA has contributed to the suppression of U.S. wages are supported only by anecdotal evidence, not serious statistical analysis. There has been no material difference in wage rates between states and industries with a large volume of U.S. imports from Mexico and those with a small volume. By far the most important channel by which Mexico influences U.S. wages is immigration and this is a function of geography, not trade. In Mexico, real wages did take a big hit in 1995 due to the peso crisis and subsequent deep recession, but rebounded sharply in the decade following 1997.

From day one, NAFTA results were skewed by the political tumult in Mexico caused by a populist revolt in Chiapas, assassinations that marred the 1994 Mexican presidential election campaign, and the peso crisis that erupted in December 1994. NAFTA obligations helped Mexico weather the financial storm of late 1994 95 in two ways: by channeling the Mexican response toward orthodox fiscal and monetary policies instead of trade protection and subsidies, and by encouraging U.S. financial support for its troubled neighbor. The Mexican economy rebounded quickly and resumed robust growth in the second half of 1996.

Since then, Mexican economic growth has been solid but not spectacular and (as of late 2007) not good enough to redress decades of economic problems. Tight fiscal and monetary policies brought the macroeconomy quickly back into balance, but constrained growth well below levels needed to solve the problems of Mexican society. Investments in physical infrastructure, human capital, and social services have been inadequate. As a result, the Mexican government has not been able to allocate the growth dividend across a broad spectrum of society, and Mexican firms and workers have not been able to take full advantage of the opportunities created by NAFTA. Moreover the growth has been highly skewed among regions within Mexico, exacerbating income disparities between the northern and southern states.

In the United States, NAFTA has had a more profound impact on trade *politics* than on trade itself. The rancorous ratification debate in the U.S. Congress in 1993 galvanized an emerging anti-globalization movement that demanded "No More NAFTAs" and a substantial restructuring of the international economic order. Recurring claims by NAFTA critics that the pact has displaced large numbers of U.S. jobs and investment have proven unfounded, however.

Nonetheless NAFTA became a lightning rod for concerns about the adverse effects of globalization on American workers and industries. The traditional protrade coalition in Congress fractured. Democrats abandoned support for new trade initiatives in response to their labor constituents and in retaliation for the highly contentious and partisan debate in the Republican-led Congress over tax policy and social issues. "Fast-track" trade negotiating authority lapsed in 1994 and major new trade initiatives were deferred until August 2002, when the Congress

narrowly restored the trade mandate under the rebranded title of "Trade Promotion Authority" (TPA). Even with TPA, however, U.S. officials struggled to secure passage of NAFTA's phonetic cousin: the U.S. Central American Dominican Republic FTA, or CAFTA-DR, passed the House of Representatives by a narrow 2-vote margin in the summer of 2005 (contrasted with a 34-vote margin for NAFTA itself in 1993).

In sum, NAFTA has had a positive effect on economic growth in all three countries, but challenges remain. Even though each country, on balance, has profited from the regional trade deal, the aggregate gains mask adjustment problems besetting members of each society. The solution to such problems for each signatory may be in domestic policies that allow its workers, farmers, and firms to take advantage of the opportunities created by NAFTA and that retrain, retool, and provide income support for those adversely affected by the new competition.

See also Central American Common Market (CACM); Central American Dominican Republic Free Trade Area (CAFTA-DR); customs unions; free trade area; Free Trade Area of the Americas (FTAA); regionalism

FURTHER READING

Hufbauer, Gary Clyde, and Jeffrey J. Schott. 2005. *NAFTA Revisited: Achievements and Challenges.* Washington, DC: Institute for International Economics. A compre hensive overview of NAFTA's first decade with detailed analyses of key sectors (autos, agriculture, energy) and issues (labor, environment, dispute settlement, migra tion) covered by the trade pact.

Martin, Philip L. 1993. *Trade and Migration: NAFTA and Agriculture.* Washington, DC: Institute for International Economics. Analyzes the "migration hump" likely to occur in response to Mexican farm reforms and NAFTA, which could increase migration flows to the United States in the short run.

Pastor, Robert A. 2001. *Toward a North American Com munity.* Washington, DC: Institute for International Economics. Assesses NAFTA's pros and cons and pro poses initiatives to deepen economic integration among the three countries.

Weintraub, Sidney, ed. 2004. *NAFTA's Impact on North America: The First Decade.* Washington, DC: Center for Strategic and International Studies. Edited volume on the economics and politics of U.S. Mexican relations during NAFTA's first decade.

World Bank. 2003. *Lessons from NAFTA for Latin America and the Caribbean Countries: A Summary of Research Findings.* Washington, DC: World Bank. In depth analysis of the impact of Mexican reforms pursuant to NAFTA on economic growth, the distribution of in come, and poverty in Mexico.

JEFFREY J. SCHOTT

◼ North-South trade

North-South trade refers to trade between rich (or developed) countries the North and poor (or developing) ones the South. The essence of North-South trade is that the two regions differ, and the variety of assumptions about the cause for instance, differences in capital per head or technology goes a long way toward explaining the variety of conclusions that are drawn about it. In addition, researchers vary as to (1) whether they consider North-South trade to mean trade that reflects only the differences between regions per se, or trade stimulated by reductions in trade barriers (natural or policy), (2) whether they permit other changes (e.g., in capital flows or technology), and (3) what they are seeking to explain for example, divergence between regions or within them (income distribution).

This entry presents a few data on North-South trade, discusses a number of models that predict divergence between North and South, and explores the hypothesis that North-South trade explains widening income inequalities.

The Data Traditionally North-South trade was conceived of as the trade of Northern manufactures for Southern primary goods, which bears a close relationship to analyses of the terms of trade under the so-called Prebisch-Singer hypothesis. This view of North-South commodity trade once had some empirical validity, but it is now severely dated.

Defining the North as the members of the Organisation for Economic Co-operation and Development as of 1990 less Turkey, and the South as all other countries, North-South trade accounted for 38.4 percent of world trade in 1962 and 38.6 percent in 2005 (provisionally), with an (oil-induced) high of 42.6 percent in 1981 and a low of 30.9 percent in 1988. One big change, however, was that whereas in 1962 the North exported twice as much to the South as it imported, the South had a trade surplus with the North over the period 1999 2005. Since 1962 the South's share of world exports to all destinations has increased from 18.1 to 40.7 percent. The story in manufacturing is similar, with a huge growth in Southern capacity and share of world trade see figure 1, which shows Southern exports growing from less than 5 percent of the world total to 37.5 percent. Within Southern exports to the North, the transformation from primary to secondary sector exports is striking (see figure 2). From a share of 10.7

Figure 1

Shares of world manufacturing trade. Source: World Bank, World Integrated Trade System (WITS).

Figure 2

Shares in Southern exports to North. Source: World Bank, World Integrated Trade System (WITS).

percent in 1962, manufactures now account for about 75 percent of exports to the North.

North-South Trade Models The mainstream neoclassical model of international trade the Heckscher-Ohlin model does not adequately explain the long-run differences in income and/or growth that are of interest in the North-South literature. It assumes identical technologies and, in the absence of complete specialization, it predicts equal factor prices in the North and South. Although assuming economywide Hicks-neutral technical differences could explain differences in average income within the Heckscher-Ohlin framework, this begs the question of why they exist. (Hicks-neutral technical differences between countries arise when the productivities of every factor production differs by the same proportion between them.) Similarly, a simple Ricardian model, in which the only difference between countries lies in the different efficiencies of labor the only factor in the various sectors, is thrown back on such a deus ex machina to explain North-South differences.

In order to explain unequal and possibly diverging incomes between North and South, researchers have had to delve into more fundamental characteristics. Marxian economists adopt the theory of "unequal exchange," whereas mainstream economists using, say, a Ricardian approach, take comparative advantage as given and locate North-South differences in, for example, the elasticities of demand for their products, their market structures, or their labor markets. The first two, frequently associated with Raúl Prebisch and Hans Singer, can explain the declining terms of trade of Southern countries and their inability to grow out of poverty. An early model of the labor-market style (Findlay 1980) had a "manufacturing" North with full employment and an "agricultural" South with fixed wages and surplus labor (as in Arthur Lewis's model of development). Growth arose spontaneously in the former and was transmitted to the latter via the demand for imports. With a fixed real wage and given technology, the critical variable in equilibrating the world economy was the terms of trade which, given unitary income elasticities of demand for both products, adjusted to
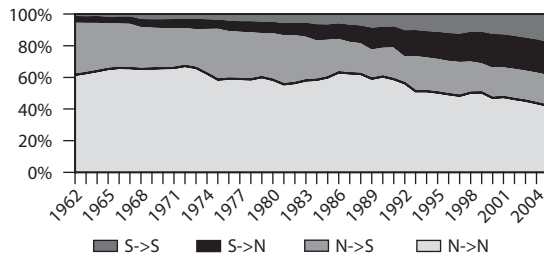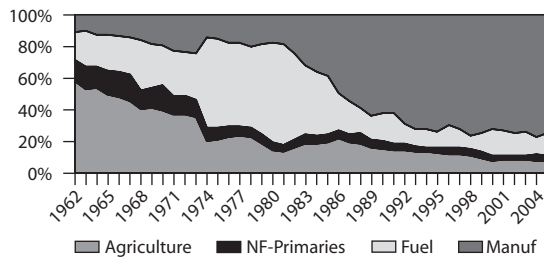
produce equal growth in the North and South even though the North had a higher income level. Increased Southern productivity depressed the terms of trade rather than boosted growth. Subsequent theoretical developments that allowed capital to move internationally or equalized profit rates in some other way produced divergence (unequal growth) even with unitary demand elasticities.

Nicholas Kaldor located the asymmetry in economies of scale of manufacturing, so that a country that had more manufacturing initially would achieve higher growth rates and be "Northern." Among several formalizations of this idea, Krugman's (1981) is the most elegant. Economies of scale in industry raise profits and accumulation in the North and eventually erode the South's industrial sector. The process stops only when one or both of the regions becomes completely specialized in production. It is likely that the South benefits from the earlier stages of this process and possible that it does so relative to autarky (i.e., no trade at all) even at the final point. And if capital is mobile between North and South as well as goods, such gains are probably more likely.

Kaldor thought of economies of scale in rather physical terms and hence as internal to the firm, but they are probably better thought of as external, residing, for example, in the thickness of labor or intermediate markets, backward or forward linkages, or spillovers of knowledge. This highlights the similarity of these models with those of economic geography. Knowledge spillovers could in fact operate throughout the economy, being related to, say, average levels of human capital, but provided that goods have different human capital intensities, trade and specialization would give the model the same feel as above. In such models human capital (skilled workers) may well have incentives to migrate from South to North.

Related to knowledge is innovation the ability to develop new products or varieties. If this skill differs between North and South this too can generate unevenness. Again Krugman (1979) offers an elegant model. Only the North can innovate but eventually the ability to produce any good becomes

worldwide as in Vernon's product cycle theory. Providing that it keeps innovating the North will employ at least some of its factors on "new" products and thus will export "new" goods in return for "old." It will earn higher incomes than the South via its quasi-monopoly in these products even though it has only equal productivity with the South in established (old) products. This implies that a reduction in the rate of innovation or an increase in the speed of imitation in the South could actually reduce real wages in the North. This introduces potential North-South conflict, a feature that some consider a necessary component of North-South trade models, although in the basic version of Krugman's model Northern innovation is always good for the South absolutely. If, however, capital were mobile, Southern workers could lose from increased innovation as capital flowed from South to North.

The simple innovation model has spawned many offspring. It has been located in richer models of production by combining it with factor abundance driven trade. Krugman's horizontal product differentiation (each new product is of the same quality as each old one) has been replaced by vertical differentiation whereby innovation creates better varieties. Innovation and imitation have been made costly and endogenized. And policies have been introduced: Northern trade restrictions typically slow down innovations at everyone's expense, except (possibly) Northern workers', and restrictions on Southern imitation (fierce intellectual property rights regimes) hurt the South and, quite possibly, the North as well.

The North-South trade literature has been mostly theoretical; it has been tested only informally against alleged stylized facts. In particular, there is no convincing empirical evidence that North-South trade has held the South back not least because it is so difficult to formulate a plausible counterfactual without such trade. There is, however, a vigorous debate about whether trade in general enhances Southern growth, income, or development. Early positive results from cross-country exercises were largely demolished by Rodriguez and Rodrik (2001). Even these authors admit, however, that there is no general presumption that openness harms average

incomes, and more careful research is starting to rehabilitate the positive view. Case studies also suggest that openness is commonly a key input to development.

A second component of empirical testing has plotted the effects of Northern research and development efforts on Southern productivity. It has suggested strong links via goods trade (perhaps via embodied technical progress), the importance of foreign direct investment, the need for local skills to ease adaptation, and the fact that knowledge apparently does not easily cross large skills gaps (e.g., Chad learns relatively little from the United States).

North-South Trade and Wages The second major stream of North-South trade research is shorter-run and empirical. It considers the effects of increasing global integration on North income inequalities over the last few decades. It stems from Wood (1994), which postulates a world of fairly free capital mobility and sector-neutral technology differences, and makes the critical difference between Northern and Southern countries their relative endowments of skilled and unskilled labor. Declining barriers to the South's exports of unskilled labor-intensive goods, particularly manufactures, were, he argued, a major explanatory factor in the widening differential between skilled and unskilled wages in the North. He showed that the unskilled labor embodied in a dollar's worth of Southern exports far exceeded that in a dollar's worth of Northern exports: the North was essentially importing labor. This drove down its price because countries produced different combinations of the many goods that existed, so that there was effectively specialization, and the factor price equalization theorem failed to apply.

No one denied the increasing skills wage gap and the weakness of unskilled wages and employment, but there were several challenges to Wood's explanation. Some argued that non-resource-based imports from developing countries were too small perhaps 3 4 percent of Northern countries' GDP to have such profound effects, especially given that the increase in labor supply embodied in net imports (i.e., imports less exports) was small. Others argued that in order to influence wages in the North,

Southern exports would also have to drive down the relative prices of unskilled-labor-intensive goods in the North, and this had not happened. In fact, however, this turned out to be only because technical progress was driving down the prices of electronic equipment, especially for information technology related uses. Among "traditional" goods, the relative price decline was evident.

The principal counter was that technical progress had a strong skill bias. Two pieces of evidence suggested this: first, the ratio of unskilled to skilled workers fell in virtually every sector and every Northern country. Under a pure factor abundance theory it should have risen as unskilled labor became cheaper (the declining aggregate demand for unskilled labor arising from the switch in the production bundle away from unskilled-intensive goods). Second, the skills wage gap widened in the South as well as in the North, whereas the factor abundance approach predicted a narrowing.

In fact, however, each piece of evidence admits other explanations and so does not overturn Wood's conclusions. The definitions of sectors in empirical work are very broad and almost always include tasks with widely different skilled/unskilled labor ratios. Thus offshoring the unskilled-intensive activities within a sector would be consistent with both competition from Southern imports and falling relative unskilled labor use in the North. Several explanations have been advanced for the growing skills wage gap in the South. The most interesting, by Feenstra and Hanson (1997), starts by arraying activities or sectors by increasing unskilled labor intensity: the South does the low-ranked ones and the North the higher-ranked ones. If Southern costs fell, the North would outsource its least skill-intensive activity to the South, where it would become the most skilled activity. In *both* countries, the relative demand for skills would increase. The cost decline could be autonomous or, as in Feenstra and Hanson, driven by flows of capital from North to South, which in turn may be responding to declining trade barriers.

There is, finally, the question of what causes technical progress: even if it is skills biased, it may be induced by the need to withstand competition from

cheaper labor in the South. This is certainly possible, but observe that except for very large changes over the longer run, the effect of technical progress on factor demands depends more on its sectoral composition than on which factor it tends to replace. If technical progress makes an unskilled-labor-intensive sector cheaper it will boost demand for it and hence for the bundle of factors it employs — that is, for an unskilled-intensive bundle. If North-South trade induces technical progress, the effects seem likely to be stronger in unskilled, not skilled, labor-intensive sectors.

In 2006, most economists accept that both autonomous skill-biased technical progress and North-South trade play significant roles in skills gap. Most accept that the latter effect is likely to get stronger in the future as trade and communications costs fall, allowing production processes to be even further divided. It is possible, however, that by making components cheaper, North-South trade sufficiently increases demand for products that require unskilled labor in the North (e.g., delivery services) that demand for Northern unskilled workers increases overall.

This literature does not deal directly with the effects of North-South trade on Southern development but is predicated on the South's strong expansion of production, as we have seen in Asia since 1960. Its consistency with rising skills wage gaps in the South is realistic and highlights one of the major challenges of future development — how to meet the (modest) income aspirations of the South's large less-skilled labor force.

See also North American Free Trade Agreement (NAFTA); primary products trade; terms of trade; trade and wages

FURTHER READING

Evans, D. 1989. "Alternative Perspectives on Trade and Development." In *Handbook of Development Economics*, vol. 2, edited by H. Chenery and T. N. Srinivasan. Amsterdam: North Holland, 1241 1304. Explores the relationship between North South models and unequal exchange.

Feenstra, R., and G. Hanson. 1997. "Foreign Direct Investment and Relative Wages: Evidence from Mexico's Maquiladoras." *Journal of International Economics* 42 (3/4): 371 94. Introduces the model in which industries locating from North to South are the least skill intensive in the former and become the most skill intensive in the latter. The result is that the average skill intensity of industry rises in both places!

Findlay, R. 1980. "The Terms of Trade and Equilibrium Growth in the World Economy." *American Economic Review* 70: 291 99. Simple model where the terms of trade bear all the strain of adjustment as the South grows.

———. 1984. "Growth and Development in Trade Models." In *Handbook of International Economics*, vol. 1, *International Trade*, edited by R. W. Jones and P. E. Kenen, Amsterdam: North Holland, 185 236. A useful survey of the early North South modeling literature.

Haskel, J. E. 2000. "Trade and Labor Approaches to Wage Inequality." *Review of International Economics* 8 (3): 397 408. Debate on skills bias versus trade in explaining Northern income inequality.

Krugman, P. 1979. "A Model of Innovation, Technology Transfer, and the World Distribution of Income." *The Journal of Political Economy* 87 (2): 253 66. Higher Northern incomes arise from innovation, which is eventually imitated in the South.

———. 1981. "Trade, Accumulation, and Uneven Development." *Journal of Development Economics* 8 (2): 149 61. A two country model with economies of scale and accumulation. An initial advantage in capital endowment confers permanent advantages on the North.

Rodriguez, F., and D. Rodrik. 2001. "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross national Evidence." *NBER Macroeconomics Annual 2000.* Cambridge, MA: MIT Press, 261 324. A strong critique of openness growth regressions.

Wood, A. 1994. *North South Trade, Employment, and Inequality: Changing Fortunes in a Skill Driven World.* New York: Oxford University Press. A detailed account of how international trade can reduce unskilled wages in the North.

World Bank. World Integrated Trade System (WITS). Available at http://wits.worldbank.org/witsnet/StartUp/ Wits Information.aspx"

L. ALAN WINTERS

offshore financial centers

Offshore financial centers (OFCs), despite their name, are not necessarily islands although many island jurisdictions have established OFCs nor do they have a single universal definition. One definition of an OFC is a jurisdiction whose financial sector accounts for a significant and disproportionate share of its domestic economy. Another definition is a jurisdiction where a majority of financial transactions conducted by its institutions are done on behalf of clients who reside in other jurisdictions.

Some of the largest offshore centers in terms of business volume are Bermuda, the Bahamas, the Cayman Islands, Jersey, Guernsey, and the Isle of Man. International financial centers with characteristics similar to OFCs include Luxembourg, Ireland, Switzerland, Singapore, and Hong Kong.

Historically, OFC jurisdictions were associated with one or more of the following: low or zero taxation, moderate financial regulation and supervision, and secrecy or anonymity in financial dealings. Increasingly, the successful OFCs are associated with niche markets in which they have developed extensive professional expertise (for example, in reinsurance, trust business, and private banking) and have reputations for maintaining good supervisory and regulatory systems. Tax advantages continue to be a key driving force for the domicile of financial transactions in OFCs. Some jurisdictions consider the term *offshore financial center* pejorative because of its historical association with lax regulatory and anti-money-laundering regimes. A number of international financial centers have many of the characteristics of offshore centers concerning tax advantages, development of instruments to attract international financial business, and cross-border financial transactions that are significantly out of proportion with the size of their domestic economy.

Growth and Activities of OFCs The growth of OFCs can be traced to the restrictive regulatory regimes in many advanced countries in the 1960s and 1970s. These regimes restricted the flow of capital to and from other countries (excluding trade financing) or imposed restrictions on the interest rates banks could offer, or raised banks' funding costs in domestic markets (for example, through the imposition of high non-interest-bearing reserve requirements). These restrictions, which in many cases were intended to provide governments with more control over monetary policy or to protect the balance of payments, encouraged a shift of deposits and borrowing to less-regulated institutions, including banks in OFCs not subject to such restrictions. As large multinationals and financial institutions shifted financial activity "offshore," the "euromarket" was established. These activities began in the financial centers of Europe (mostly London) and soon spread to other offshore centers.

As with international financial centers, business conducted in OFCs covers a wide range of financial sectors such as banking, insurance, securities, and some nonfinancial activities such as shipping registries and company and trust service providers. Most OFCs specialize in specific types of financial services, however. Multinational corporations and

high-net-worth persons are some of the most frequent users of OFCs.

Banking is the most prevalent business, and a significant volume of international banking transactions are conducted in OFCs. Most banks located in OFCs are branches or subsidiaries of international banks. Their main activity is collecting deposits from various markets and channeling them back to their parent institutions. Private banking is a major service offered to high-net-worth persons. Specialized services for such clients range from asset management to estate planning, foreign exchange trading, and pension arrangements.

Collective investment schemes (mutual funds and hedge funds) are also domiciled in OFCs, mainly for tax purposes. Related fund activities such as fund distribution, asset management, fund administration, custodian services, and back-office work are also conducted in these centers. Although a large number of hedge funds are domiciled and administered in OFCs, the management of the investment portfolios is normally conducted in the major financial centers such as London and New York.

Insurance business, including life, property, reinsurance (an insurance company that assumes all or part of a risk undertaken by another insurance company), and captive (a company owned by a noninsurance firm that provides insurance coverage to the owner), is also conducted in some OFCs. Innovative regulatory and legal environments have helped OFCs attract a large share of the world's captive and reinsurance market. For example, most of the new capital entering the reinsurance industry on the year following the catastrophic hurricane season of 2005 was placed in existing or start-up companies in OFCs. A large portion of the world's captive insurance companies is also domiciled in these centers.

A significant number of special purpose vehicles (SPVs) that is, financial instruments that are tailored to meet the particular financing needs of financial and nonfinancial corporations are registered in OFCs. Financial firms use SPVs for securitization, and nonfinancial corporations use them to lower the cost of raising capital. OFCs are attractive places to register SPVs because of the tax advantages they offer.

Asset protection, including the establishment of trusts, is another service offered by OFCs. Reasons for managing assets in OFCs include protection from weak domestic banks or currencies, additional legal protection from lawsuits in the home jurisdictions, and tax efficiency/avoidance.

Concerns about OFCs Some of the characteristics of OFCs raise concerns about potential risks to the international financial system. First, since the livelihood of OFCs depends on their ability to attract global financial business, competition among jurisdictions is strong. Such competition is beneficial when it contributes to innovation in financial instruments and products and lowers costs of financial services worldwide. It can also raise concerns, however, if the lower cost of financial services is achieved by lowering regulatory and supervision standards. Second, as OFCs provide financial services predominantly to nonresidents, the authorities in the transaction home countries are interested in the impact on their national economies of the operations in OFCs, especially when these operations are beyond the home country authorities' control. Third, anonymity and lack of transparency in OFCs, opaqueness of the operations of offshore corporations, and legal protections in some OFCs have created concerns. Historically, a lack of reliable data on activities in OFCs hampered analyses of the potential risk that OFCs pose to financial stability. SPVs domiciled in OFCs have been implicated in some of the large corporate frauds. Fourth, the anonymity of financial transactions makes OFCs attractive to money launderers. OFCs are a potential gateway for the proceeds of crime to gain access to and to be laundered through global financial markets. Money laundering is an international problem, and to be effective, international standards against money laundering and terrorist financing need to be applied worldwide, especially in large financial centers.

Concerns about the potential risk posed by OFCs resulted in OFCs being placed on "name and shame" lists by a number of international bodies in 2000, including the Financial Stability Forum (FSF), the

Financial Action Task Force on Money Laundering (FATF), and the Organisation for Economic Cooperation and Development (OECD). This posed significant reputation damage to OFCs. Also in 2000, the International Monetary Fund (IMF) was asked to intensify its oversight of OFCs in the context of its mandate to promote financial stability. The IMF's program has focused on (1) regular monitoring, (2) assessing OFCs' compliance with international supervisory and anti-money-laundering standards, (3) enhancing the transparency of activities in OFCs, and (4) strengthening international cooperation and information on supervisory and regulatory systems.

Because of the intensified international attention, many OFCs upgraded their supervisory and regulatory systems and improved the transparency in their operations. A number of smaller OFCs also decided to exit from part or all of their OFC activity after weighing the costs and benefits. Although a financial center can be a useful addition to an economy, a significant and costly investment in the supervisory and regulatory infrastructure is needed to provide the internationally accepted minimum supervisory system required to avoid reputation risk. At the same time, a number of new and aspiring OFCs are being established where the authorities have been willing and able to invest heavily in the necessary infrastructure.

See also capital mobility; hedge funds; international financial centers; International Monetary Fund (IMF); money laundering; trade in services

FURTHER READINGS

Darbar, Salim, et al. 2003. "Assessing Offshore Financial Centers Filling a Gap in Global Surveillance." *Finance and Development* 40 (3) (September). Available at www.imf.org/external/pubs/ft/fandd/2003/09/pdf/darbar .pdf.

International Monetary Fund. http://www.imf.org/external/ np/mae/oshore/2000/eng/back.htm. IMF OFC Web site that reviews the issues raised by OFCs and the IMF's program to assess OFCs against international regulatory standards.

R. BARRY JOHNSTON

■ **oil**

See Organization of the Petroleum Exporting Countries (OPEC); petroleum

■ **oligopoly**

Oligopoly is a situation in which several firms compete over a given market, but there are so few firms that each one can earn a profit over and above what is earned in the more competitive industries. Barriers to entry in the form of information costs, economies of scale, franchises, and patents appear to be sufficient to keep out enough new competitors. Thus it is convenient here to think of oligopoly as imperfect competition without free entry, and monopolistic competition as imperfect competition with free entry.

Not surprisingly, international trade economists concerned with oligopoly have focused on the role of oligopoly profits. With monopolistic competition or perfect competition, there are no profits measured as the return over all opportunity costs. The existence of economic profits raises three large questions: How do those profits affect the pattern and volume of international trade? How do those profits affect who gains and who loses from international trade? And do those profits present policymakers with trade policy opportunities?

There are two approaches to studying oligopoly and trade: partial equilibrium and general equilibrium. The partial equilibrium approach concentrates on the world industry with no attempt to explain the division between exporters and import-competing firms or how oligopoly profits affect demand. This approach is most useful in studying questions involving trade policy. The general equilibrium approach seeks to examine how oligopoly profits affect the traditional trade questions of determining what is traded and the volume of trade, and identifying the winners and losers.

The Partial Equilibrium Approach In the first stage of development, the partial equilibrium approach dominated the literature with representative contributions by Dixit (1984), Brander and Spencer (1985), Eaton and Grossman (1986), and Brander

and Krugman (1983). Global firms can sell in an integrated world market, segmented markets in each country (Dixit 1984; Brander and Krugman 1983), or in some outside country (Brander and Spencer 1985; Eaton and Grossman 1986). The market is integrated if arbitrage keeps prices at home tied to foreign prices. The market is segmented if there is no arbitrage to keep prices in different markets in line. With tariffs and quotas, imperfect markets are best considered segmented.

Probably the most interesting and yet most controversial contribution is the one by Brander and Spencer (1985), which shows that subsidizing the exports from a domestic firm can be beneficial to the economy by a strategic trade policy that transfers oligopoly profits from a foreign firm to the home firm in the same industry. This result was made intuitively clear by Eaton and Grossman (1986), who showed that by a different assumption regarding the behavior of firms it was optimal to tax rather than subsidize the export of the oligopoly. Why the difference? The explanation is extremely simple: if we assume the oligopolist is too conservative toward rivals, the government should subsidize its exports; if the oligopolist is too aggressive, the government should tax its exports. If the economist cannot distinguish between these two cases in practice, then our information does not allow a recommendation. The difference arises in the model as the difference between Cournot and Bertrand competition. Cournot is competition in quantities; Bertrand is competition in prices. Assume a firm in country A and one in country B compete in a third market with goods that are imperfect substitutes. In the case of Cournot competition, the firm assumes that the other firm will maintain its output; in truth, if it expands output, the other will contract its output. Had the firm known this, it would have expanded its output even more. In the case of Bertrand competition, the firm assumes the other firm will maintain its price; in truth, if it lowers the price, the other firm will lower its price. Had the firm known that, it would have lowered its price by a smaller amount. Thus the existence of oligopoly profits does not seem to have raised the likelihood of a beneficial trade policy over and above traditional optimum tariff arguments. Even if strategic trade policies can be identified, they run into the same familiar problems: (1) governments correctly calculating optimizing tariffs or subsidies, (2) foreign retaliation can eliminate the benefits, (3) costly economic rent seeking by industries alerted to the policies, and (4) raising the cost of classifying goods for customs enforcement.

Brander (1981) and Brander and Krugman (1983) examine the role of price discrimination in international trade. In this case, it is assumed that there is an international oligopoly in which each country is a segmented market. Oligopolistic firms can raise profits by cutting prices in the market with a higher elasticity of demand. This implies that the standard theory of price discrimination can be applied, because now firms, regardless of their location, will sell in all markets at the prices prevailing in each segmented market. There now may be cross-hauling, or wasted transportation costs as even identical goods move in both directions (Brander and Krugman 1983). A felicitous mental picture would be two logging trucks going in opposite directions and carrying identical logs.

General Equilibrium and Oligopoly In the general equilibrium approach, it is important to explain the pattern of exports and imports as well as the costs and benefits of trade. The first problem to overcome is how to embed oligopoly in a general equilibrium model that was designed with perfect competition in mind. Some theorists have stressed the difficulties entailed by considering large firms, such as the possibilities of monopsony (or buyer power) in factor markets or considerations about what the large firm should maximize. Following the seminal paper by Lerner (1934) for the case of monopoly, Markusen (1981), Neary (2003), and Ruffin (2003a, 2003b) assume that it is more convenient to consider firms to be large in their industry but small in the economy. In this case, the firm can treat all factor prices and prices in other industries as fixed. Ruffin (2003a) shows that the firm's profit functions are then independent of the numéraire selected, so the problem of what the firm maximizes is avoided. In this case, clear and simple results can be reached.

Both Neary (2003) and Ruffin (2003a, 2003b) also employ the assumption that firms act as Cournot oligopolists; that is, they assume the outputs of their rivals are constant. Neary (2003) takes the assumption literally that firms are small in the economy by considering a continuum of firms, whereas Markusen (1981) and Ruffin (2003a, 2003b) simply consider a finite number of firms acting as if they are small in the economy.

The general equilibrium approach is most easily exhibited in a model in which Cournot oligopolies are selling in an integrated world market. Again, it is oligopoly profits that provide the key. How does oligopoly affect the costs and benefits of international trade? Consider two countries, Home and Foreign, that are the same size, with each having a comparative advantage in a good facing the same demand and costs. Since trade economists already know the impact of different factor intensities across industries, a Ricardian model (Neary 2003; Ruffin 2003a) with only one factor of production is often used. We shall call that factor "labor" but it could be a bundle of resources used in the same proportions in all industries. In this case, if we assume constant returns to scale, the production-possibility frontier for the economy is linear. With these assumptions we can assume that both countries have the same wages under perfect competition. Moreover, if we assume that under oligopoly all industries have the same degree of

competition (same number of firms), it will also be true that both countries will have the same wages with or without trade. Thus if we take wages to be fixed per unit labor (as the unit of measure), we can take dollar costs as given for each country. Thus suppose, as in table 1, that it costs Home $4 to produce a bushel of apples and $6 to produce a bunch of bananas, and it costs Foreign $6 to produce apples and $4 to produce bananas. Note that it is best to take wages as the unit of measure rather than one of the commodities, because in imperfect competition firms are choosing commodity prices to maximize profits.

It will be useful to summarize the Cournot oligopoly model when the product is homogeneous. The key assumption is that firms conjecture that rivals will maintain their output, a conjecture that holds in the Cournot-Nash equilibrium. There is a Cournot-Nash equilibrium if no firm can make an additional profit by changing its strategy, given the strategies of all of the other firms. If $D(P)$ is the demand for the homogeneous product, $D'(P)$ the slope of the demand curve, x_i the output of the ith firm, and c_i the marginal cost of the ith firm, then the Cournot equilibrium is:

$$D'(P)(P - c_i) + x_i = 0 \qquad (1)$$
$$\Sigma x_i = D(P) \qquad (2)$$

If there are N firms, this is a system of $N+1$ equations. The demand functions clearly must reflect all

Table 1
Gains to workers under competition and oligopoly

State	Country	Autarky		Free trade		Gains to workers
		apples	bananas	apples	bananas	
Perfect competition						
	Home	$4	$6	$4	$4	
	Foreign	$6	$4	$4	$4	17%
Tight oligopoly (N 2)						
	Home	$8	$12	$6.67	$6.67	
	Foreign	$12	$8	$6.67	$6.67	31%
Loose oligopoly (N 3)						
	Home	$6	$9	$6	$6	17%
	Foreign	$9	$6	$6	$6	

prices and incomes. Ruffin (2003b) describes the general equilibrium equations, taking account of the incomes of all factors of production and assuming all agents have identical, homothetic utility functions. Equations (1) and (2) replicate rivals holding output constant because when the ith firm adjusts its output, $D'(P)$ is the change in its demand by the resulting change in price. The interpretation of (1) is that if the firm adjusts output by $D'(P)$ by causing price to rise by a dollar, it loses the profit margin on those units but gains x_i on the resulting sales. But this becomes greatly simplified even in the case of nonidentical firms by adding the profit-maximizing equations (1) over the N firms and defining the price elasticity of demand as $\varepsilon = -D'(P)P/D$, resulting in:

$$P = \Sigma c_i/(N - 1/\varepsilon) \qquad (3)$$

For simplicity, suppose that $\varepsilon = 1$, which works for Cournot oligopoly for $2 \leq N$ (Ruffin 1971). Since oligopolists price according to elasticities of demand, this equation can be applied in general equilibrium without worrying about the impact of oligopoly profits on demand as long as we assume world symmetry so that costs can be taken to be independent. In a Ricardian trade model in which both countries have the same income and face the same demand for the good in which they have a comparative advantage, equation (3) works well because we merely have to add the costs of all the firms active in both countries and divide by 1 less than the number of firms in each world industry.

Consider first perfect competition in which prices must equal marginal costs. Table 1 shows both autarky and free trade. Thus in autarky, apples will be $4 and bananas $6 in Home, and the opposite in Foreign. Now if international trade is opened, in a competitive situation only Home would produce apples and only Foreign would produce bananas. Thus people of the world would pay only $4 for both apples and bananas. The gains from trade are simple to measure in this case, since equal amounts of money are spent on apples and bananas: The price of the importable falls by 33 percent, and the gain from trade for the economy as well as workers (the same in

this case) would be roughly one-half of 33 percent, or about 17 percent.

Under oligopoly both workers and oligopolists share in national income. There are two situations: a tight oligopoly in which the oligopoly profits protect high-cost producers from being driven out of business and a loose oligopoly in which oligopoly profits are insufficient for high-cost firms to survive the competition from low-cost rivals. For the moment assume the same degree of oligopoly in both industries, with only 2 or 3 firms in each country producing apples or bananas. With trade, each world industry has the potential of 4 or 6 firms. In the example, a tight oligopoly is when each country has only 2 firms producing apples and bananas, and a loose oligopoly prevails when each country has 3 firms producing both goods. Applying formula 3 to the autarkic state, as illustrated in table 1, we can see that under tight oligopoly Home will sell apples for $8 and bananas for $12. When trade is opened, apples will now be produced by 2 firms with costs of $4 and 2 firms with costs of $6, so according to formula 3 the price of apples (and bananas) will be $6.67 [= {2($4) + 2($6)}/3 = $20/3]. Economic logic tells us that the price must be less than $8, because there are more firms. For Home (and Foreign is symmetrical), the price of apples falls by 1.33/8, or 17 percent, and bananas by 5.33/12, or 44 percent, for a weighted average of 31 percent. Since we can hold worker wages the same (as the numéraire), workers must gain 31 percent from the opening of trade, a substantial improvement over the case of perfect competition. In this case, with equal degrees of competition between the two sectors, the gains from trade to the *economy* will be smaller than in the case of perfect competition because the inefficient banana industry in Home and apple industry in Foreign still operate. Accordingly, since workers gain more than under perfect competition, it follows that trade causes substantial losses to the oligopolists compared with autarky. *Nevertheless, the remaining oligopoly profits of the lower-cost firms serve to protect the higher-cost firms from competition, just as tariffs do.*

If we have looser oligopolies, the situation may be quite different. As shown in table 1 again, suppose

that the number of firms in each industry prior to trade is 3. Then in Home, applying equation (3) once again shows that under autarky apples will sell for $6 and bananas for $9, with the opposite in Foreign. When trade opens, the Foreign apple industry and the Home banana industry will be unable to compete with their low-cost counterparts. Indeed, under Cournot, they will assume that any output will lower price below $6, so that the equilibrium price of each good will be $6. Notice that in this case we get a result similar to perfect competition: trade causes the shutdown of inefficient industries, and both oligopolists and workers gain precisely the same amount as under perfect competition.

The looser oligopoly drives higher-cost firms in the other country out of business simply because price is closer to their marginal costs. This is also a measure of the robustness of a country's comparative advantage. The more robust the comparative advantage, in the sense of the greater the difference in costs between the Home and Foreign countries, more oligopoly power is compatible with driving foreign rivals out of business. Had we assumed $4 costs for apples in Home and $8 costs for apples in Foreign, only two Home firms would have been sufficient to drive the high-cost firms out of business.

In the tight oligopoly case examined in table 1, both high-cost and low-cost firms share the market. With perfect competition or a loose oligopoly, the high-cost firms are driven out of business. In general, this means that oligopoly power can reduce the degree of specialization and so reduce the volume of international trade. This helps to explain one of the most interesting observations about world trade in the last 50 years, namely that it has increased at about twice the rate of world output. People have usually attributed this to reductions in communication costs, transportation costs, and trade barriers. But the deregulation of industry and the rise of domestic competition may be another cause.

The Gains from Trade We assumed in table 1 that there are equal degrees of oligopoly in both sectors of the economy. This implies that the gains from trade in the economy cannot exceed what would prevail under perfect competition, although

workers can clearly gain more because of the procompetitive effects of trade on both export and import prices. But some authors have argued that with oligopoly the trade gains to the economy are greater than under perfect competition! As discussed in Ruffin (2003a, 2003b), this may arise when there are different degrees of competition between sectors. Markusen (1981) considers a two-sector model with perfect competition in one sector and monopoly in the other, and assumes that when international trade opens there is a two-firm Cournot oligopoly of Home and Foreign monopolists. He shows that even in the absence of actual trade, the potential for trade can improve welfare due to the procompetitive effects of the resulting two-firm world oligopoly. Markusen proves the result by making the strong assumption that under perfect competition there are no comparative advantages and, hence, no gains from trade. The existence of monopoly or oligopoly in part of the economy then improves the gains from trade because opening markets reduces those oligopoly powers. But there is no general rule that under oligopoly or monopoly the gains from trade are greater than under perfect competition, as pointed out by Ruffin (2003a). In fact, Ruffin (2003b) gives an example in which the country exporting the good produced by an oligopoly gains more from trade than under perfect competition, but the country importing that good gains less. The reason has to do with the unfavorable terms of trade effects related to exporting goods produced under competition compared with oligopoly.

Oligopoly also has important implications for the effect of trade on income distribution. Standard Heckscher-Ohlin trade theory predicts that abundant factors of production gain from trade while scarce factors lose from trade. The reason for this is that trade shifts resources from industries intensive in scarce factors to industries intensive in abundant factors. The imbalance of supply and demand requires the prices of abundant factors to rise, and those of scarce factors to fall. Thus a key characteristic of Heckscher-Ohlin theory is that the opening of international trade, or the expansion of trade from lower trade barriers, should cause a factor of

production to lose in one country and gain in another country. The existence of oligopoly moderates this effect, and it is possible for oligopoly profits to rise everywhere or fall everywhere. In the case of loose oligopolies, trade can cause profits to soar because low-cost firms drive out high-cost firms and their gain in profits from the world market will offset the loss of profits from the high-cost firms. This can happen everywhere. With tight oligopolies, low-cost firms must compete with high-cost firms and their profits will fall everywhere.

The case of workers is quite different. It is possible to show that, abstracting from Stolper-Samuelson effects arising from different factor intensities, workers must always gain from international trade. This holds under quite general circumstances, regardless of differences in the degree of competition across the economy. The reason is that under all circumstances, prices must fall measured in terms of wage units. Examining equation (3) shows this. Unless the price elasticity of demand falls (an unlikely event in a larger market), an increase in the number of firms (both low- and high-cost) will cause the price of the product to fall for both export goods and import-competing goods. If there is the same number of firms, but the high-cost firms disappear compared with autarky, the prices of imported goods fall while the prices of exported goods remain the same: workers still gain. This, of course, ignores the adjustment costs that workers will suffer moving from high-cost to low-cost industries, as would be the case for perfect competition as well. So oligopoly raises no new issues here except for the possibility of long-term contracts that may be present in concentrated industries that may not be present in more competitive industries. The mobility of workers under oligopoly is an issue that can and should be studied.

The importance of oligopoly in international trade focuses on the role of profits. Trade policy that shifts profit from one country to the other faces informational problems and the possibility of retaliation. Generally speaking, oligopoly makes it more likely that trade will benefit nonprofit factors such as workers. Oligopoly can make trade gains to the economy smaller if oligopoly is widespread, but only if oligopoly is concentrated in the export sectors can trade be more beneficial to the economy.

See also comparative advantage; foreign direct investment under oligopoly; gains from trade; Heckscher-Ohlin model; monopolistic competition; New Trade Theory; Ricardian model

FURTHER READING

Brander, James A. 1981. "Intra industry Trade in Identical Commodities." *Journal of International Economics* 11 (1): 1 14. Assumes segmented markets in each country so that oligopoly firms in separate countries sell in each market even if the commodity is homogeneous.

Brander, James A., and Paul Krugman. 1983. "A 'Reciprocal Dumping' Model of International Trade." *Journal of International Economics* 15 (3 4): 313 21. Price discrimination applied to international trade.

Brander, James A., and Barbara J. Spencer. 1985. "Export Subsidies and International Market Share Rivalry." *Journal of International Economics* 18 (1 2): 83 100. Shows an export subsidy to a home firm that competes with a foreign firm in a third market is optimal for the case of Cournot competition.

Dixit, Avinash K. 1984. "International Trade Policy for Oligopolistic Industries." *Economic Journal* 94 (1): 1 16. A survey of partial equilibrium models.

Eaton, Jonathan, and Gene M. Grossman. 1986. "Optimal Trade and Industrial Policy under Oligopoly." *Quarterly Journal of Economics* 101 (2): 383 406. Shows an export duty on a home firm that competes with a foreign firm in a third market is optimal for the case of Bertrand competition.

Krugman, Paul R. 1979. "Increasing Returns, Monopolistic Competition, and International Trade." *Journal of International Economics* 9 (4): 469 79. A simplification of the Dixit Stiglitz model.

Lerner, Abba. 1934. "The Concept of Monopoly and the Measurement of Monopoly Power." *Review of Economic Studies* 1 (3): 157 75. Explains the economic inefficiency of monopoly in a general equilibrium context.

Markusen, James R. 1981. "Trade and Gains from Trade with Imperfect Competition." *Journal of International Economics* 11 (4): 531 51. The existence of monopoly can increase the gains from trade under some circumstances.

McGahan, Anita M., and Michael E. Porter. 1999. "The Persistence of Shocks to Profitability." *Review of Economics and Statistics* 81 (1): 143 53. Shows that industry characteristics determine the persistence of oligopoly profits.

Neary, J. P. 2003. "The Road Less Traveled: Oligopoly and Competition Policy in General Equilibrium." In *Economics for an Imperfect World: Essays in Honour of Joseph Stiglitz*, edited by R. Arnott, B. Greenwald, R. Kanpur, and B. Nalebuff. Cambridge, MA: MIT Press, 485 500. Surveys various approaches to oligopoly in general equilibrium and introduces an approach based on Cournot oligopoly with a continuum of firms.

Ruffin, Roy J. 1971. "Cournot Oligopoly and Competitive Behavior." *Review of Economic Studies* 38 (4): 493 502. Clarifies the relationship between oligopoly, perfect competition, and the number of firms.

. 2003a. "Oligopoly and Trade: What, How, and For Whom?" *Journal of International Economics* 60 (2): 315 35. Shows trade gains under oligopoly can be larger or smaller than under perfect competition in a Ricardo Cournot setting.

. 2003b. "International Trade under Oligopoly Conditions." *Review of International Economics* 11 (4): 577 87. Gives some simple examples of oligopoly and trade as well as describing the general equations of an oligopolistic world economy.

ROY J. RUFFIN

■ optimum currency area (OCA) theory

Optimum currency area (OCA) theory originates from two seminal articles in the early 1960s by the economists Mundell (1961) and McKinnon (1963). These articles drew on contemporary debates about fixed versus flexible exchange rates, treating a common currency as the extreme case of a fixed exchange rate. At the time, under the Bretton Woods system of fixed exchange rates, the choice of exchange rate was seen as a theoretical rather than practical issue, and applicability of the OCA analysis to the choice of currency domain was limited by the almost universal one country one currency rule (Pomfret 2005). The shift to generalized floating of the major currencies, however, and the move to European Monetary Union (EMU), which began in the early 1970s, generated practical interest in currency areas that has continued to grow.

Despite the huge subsequent literature and increased practical applicability, the theory of currency areas has remained centered on the Mundell-McKinnon approach of identifying the area within which the macroeconomic flexibility gains from having an independent currency are significant. This article focuses on OCA theory and the empirical evidence, while the complementary article on common currency analyzes actual experience with shared currencies.

Variants of OCA Theory In OCA theory the key issue is the extent to which macroeconomic policy can be effective in an open economy. An external shock, such as a recession induced by a drop in export demand, can be countered by cutting interest rates to stimulate investment. An economy with capital mobility and a fixed exchange rate does not have monetary policy independence, however, because the cut in interest rates will lead to capital outflows and an unsustainable balance of payments deficit. Mundell (1961) identified the point at which currency independence becomes worthwhile as the point at which factors cease to be mobile. The boundary of the OCA is set by breaks in factor mobility, that is, points at which the movement of labor or capital becomes more difficult.

The openness of an economy also can determine the effectiveness of exchange rate changes as a macroeconomic policy instrument. In very open economies, currency devaluation is ineffective because prices and wages immediately increase to remove any competitive advantage. Openness undermines the money illusion (or failure to distinguish nominal from real price changes) that permits the exchange rate to be an effective policy instrument. Larger economies have lower ratios of trade to gross domestic product (GDP), and currency devaluation can have a positive impact on output. Thus very open economies are suboptimal currency areas, but for a larger and less open economy the effectiveness of the exchange rate as a policy tool increases;

at some point the currency area becomes optimal (McKinnon 1963).

In the decades following the original contributions of Mundell and McKinnon, subsequent writers lengthened the list of criteria that might be relevant to an assessment of the optimality of a currency area. Recent contributions have argued that the trade-off may be mediated by history and by geography (Alesina and Barro 2002). With multiple criteria a unique ranking of suitability for currency union is unlikely to exist and the theory becomes difficult to test, but the essential structure of OCA theory has remained remarkably robust since 1961.

OCA theory characterizes the choice of currency area as a cost-benefit analysis trading off microeconomic efficiency against macroeconomic flexibility (Krugman 1993). Microeconomic efficiency would be maximized with a global currency. Thus, subglobal OCAs imply the existence of distortions, such as nonclearing labor markets that lead to involuntary unemployment, whose negative effects can be reduced by macroeconomic policy. Another benefit of a larger currency area is that disturbances are likely to be offsetting, so that exchange rate changes are smaller, with less feedback on domestic prices. The greater price stability is usually ascribed to random shocks being offsetting, but other mechanisms include reduction in the impact of outliers in the consumer price index, or CPI (i.e., goods with a particularly large weight in a region's consumption bundle), and reduction in the ratio of trade (or rather transactions denominated in potentially volatile foreign currencies) to GDP.

The cost-benefit perspective suggests an equal role for microeconomic and macroeconomic analysis, but the currency area literature has been dominated by macroeconomists, in part because the micro benefits of reduced transactions costs from a common currency are simple to visualize but difficult to measure. Although the benefit from a common currency (or fixed exchange rate) in terms of lower transactions costs has long been accepted as the overwhelming argument for mini states (such as Luxembourg or Brunei) not to have independent currencies, the argument becomes less potent as the currency area becomes large enough to have well-functioning foreign exchange (including forward) markets. Thus the increasing efficiency of financial markets becomes an argument for OCAs becoming smaller because it reduces the benefits via reduced transactions costs of a common currency. This may explain why it is feasible to have a small currency domain, as in countries such as Iceland or New Zealand, without incurring huge costs, but whether the trade-off between macropolicy flexibility and transactions costs is optimal for these small states is another question, and one which is impossible to answer without a good understanding of the nature and size of the transactions costs associated with independent currencies. Before the beginning of the 21st century little research had been done on the transactions cost benefits of a common currency, but this has changed since the controversial article by the economist Andrew Rose was published in 2000.

Testing OCA Theory The OCA literature has dominated the theoretical explanation of currency areas for almost a half century. There have, however, been few systematic tests of OCA theory and little positive support for OCA theory as a useful way of explaining the composition of existing currency areas or of predicting changes in currency domains. One empirical challenge was the rarity of changes in currency areas during the 1970s and 1980s. In the early 1990s it was argued that the stability of actual currency arrangements may be explained by switching costs (Dowd and Greenaway 1993), but the many changes in currency arrangements in Eastern Europe in the 1990s and the introduction of the euro in Western Europe suggest that the mechanics of changing currency arrangements are neither difficult nor especially costly.

With multiple OCA criteria, the theory becomes difficult to test: a small open economy has the biggest potential gain from joining a larger currency area in order to reduce transactions costs, but it may also be most vulnerable to external shocks and hence has the most to lose from giving up the exchange rate as a macropolicy instrument. The economists Kreinin and Heller (1974) synthesized the various criteria into the single question of whether a country could

better deal with external imbalance through devaluation or through adjustment of domestic demand. Their conclusion was that Italy, Sweden, and Switzerland were the three Organisation for Economic Co-operation and Development countries most likely to abandon their national currencies. Thirty years later only one of the three had done so, while ten of the "less likely" countries had abandoned their national currencies.

If currency areas become "optimal" ex post, then OCA theory may be untestable. A common currency might promote closer trade links and more synchronized cycles, both of which are OCA criteria; closer trade ties increase the benefit from a common currency, and synchronized cycles reduce the cost of giving up independent currencies. This is not a theoretical result but a hypothesis to be tested empirically because more bilateral trade could promote interindustry specialization and less synchronized cycles. Using various measures of bilateral trade intensity and cycle synchronization for 21 developed economies, Frankel and Rose (1998) find a robust relationship between the two variables, which they interpret as evidence that as a common currency promotes bilateral trade it also increases cycle synchronization. Thus actual currency areas fit OCA criteria better than potential currency unions, and the OCA criteria are endogenous.

The impact of currency union on bilateral trade flows has been the subject of a burgeoning literature initiated by the economist Rose (2000). Using a gravity model, Rose found that currency union has a large effect on bilateral trade, which he interpreted as evidence that a common currency substantially reduces transactions costs. Although it is plausible that a common currency reduces transactions costs and stimulates trade, the magnitude of the common currency effect is hotly debated. In Rose's study the countries in currency unions are not from a random draw; several authors have shown that currency union members are smaller and more open than their natural comparators and that history (usually in the form of colonial background) matters.

The debate has been conducted with analysis of currency area changes over time. Analyzing time-series data for correlations between changing currency union status and bilateral trade flows, Glick and Rose (2002) estimated that dissolution of a currency union halves bilateral trade. Currency union breakup is, however, usually associated with other events that disrupt trade; out of some 60 cases of post-1947 currency union dissolutions in the Glick-Rose data set, more than two-thirds broke up within a decade of the end of a colonial relationship, and the end of the ruble zone, which is not in the data set, would increase the percentage still further. In tranquil currency union changes, notably Ireland's secession from its currency union with the United Kingdom in 1979 and subsequent participation in the process leading to the euro, the impact on bilateral trade is unclear. The weaker the link between currency union and bilateral trade, the less convincing is the claim that OCA criteria become self-fulfilling ex post.

In practice, irrespective of whether the criteria may become endogenous, most of the literature on currency domains treats OCA theory as having predictive capability. Yet, the general track record of OCA theory in explaining the monetary history of the post-1945 international economic order has been miserable. Despite the increasing openness of national economies and increasing capital mobility, both unambiguous pressures for larger currency areas according to OCA theory, the number of currencies has increased substantially and the geographical size of currency domains has shrunk correspondingly. The sole significant exception is the introduction of the euro, but in Europe and the former Soviet Union as a whole there were more currencies in 2002 than a decade earlier. Globally, over the last half century, the exogenous increase in the number of countries drove the number and size of currency areas, and the OCA criteria were irrelevant to explaining this pattern. In sum, although OCA theory has dominated the analysis of currency domains, the empirical support for the theory is weak.

See also Bretton Woods system; capital mobility; common currency; dollar standard; dominant currency; euro; European Monetary Union; exchange rate regimes; impossible trinity; multiple currencies

FURTHER READING

Alesina, Alberto, and Robert J. Barro. 2002. "Currency Unions." *Quarterly Journal of Economics* (May): 409 36. A recent restatement of OCA theory.

De Grauwe, Paul. 2000. *The Economics of Monetary Union*. 4th rev. ed. Oxford: Oxford University Press. The best textbook treatment of monetary union.

Dowd, Kevin, and David Greenaway. 1993. "Currency Competition, Network Externalities, and Switching Costs: Towards an Alternative View of Optimum Currency Areas." *Economic Journal* 103: 1180 89. An influential explanation of the costs of changing currency arrangements, although to some extent countered by the relative lack of technical problems associated with the introduction of new currencies in the former Soviet Union and the introduction of the euro.

Frankel, Jeffrey, and Andrew Rose. 1998. "The Endogeneity of the Optimum Currency Area Criteria." *Economic Journal* 108 (July): 1009 25. An attempt to introduce systematic cross country evidence into the debate about OCA criteria.

Glick, Reuven, and Andrew Rose. 2002. "Does a Currency Union Affect Trade? The Time Series Evidence." *European Economic Review* 46 (June): 1125 51. A follow up to Rose's 2002 paper, which used cross country data.

Kreinin, Mordechai, and H. Robert Heller. 1974. "Adjustment Costs, Optimal Currency Areas, and International Reserves." In *International Trade and Finance: Essays in Honour of Jan Tinbergen*, edited by Willy Sellekaerts. London: Macmillan, 127 40. A pioneering attempt to introduce empirical evidence into the debate over the OCA criteria, although limited to European countries.

Krugman, Paul. 1993. "What Do We Need to Know about the International Monetary System?" *Essays in International Economics No. 190*. International Economics Section. Princeton, NJ: Princeton University (July). An important reminder not to neglect transactions costs and to remember the microeconomic benefits from a well functioning international monetary system (and a common currency).

McKinnon, Ronald. 1963. "Optimum Currency Areas." *American Economic Review* 53: 717 25. One of the two articles initiating OCA theory.

Mundell, Robert. 1961. "Theory of Optimum Currency Areas." *American Economic Review* 51: 657 65. The original OCA theory.

Pomfret, Richard. 2005. "Currency Areas in Theory and Practice." *Economic Record* 81 (253): 166 76. A recent review of the evidence on OCA theory.

Rose, Andrew. 2000. "One Money, One Market: The Effect of Common Currencies on Trade." *Economic Policy* 30 (April): 9 45. An innovative and controversial cross country study of the impact of a common currency on bilateral trade flows, and implicitly on transactions costs.

Tower, Edward, and Thomas Willett. 1976. "The Theory of Optimum Currency Areas and Exchange Rate Flexibility." *Princeton Special Papers in International Economics, No. 11*. International Finance Section. Princeton, NJ: Princeton University (May). A classic survey of OCA theory.

RICHARD POMFRET

■ Organisation for Economic Co-operation and Development (OECD)

The Organisation for Economic Co-operation and Development (OECD) is an international organization located in Paris. It was founded in 1960 as the successor to the Organisation for European Economic Co-operation (OEEC), which was established in 1947 to administer American and Canadian aid under the Marshall Plan for reconstruction of Europe after World War II. OECD members are governments of nation-states that meet the criteria of pluralistic democracy and free market economies. The founding members were Austria, Belgium, Canada, Denmark, France, Germany (Federal Republic), Greece, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States. Since then 10 more countries have joined: Australia (1971), Czech Republic (1995), Finland (1969), Hungary (1996), Japan (1964), Korea (1996), Mexico (1994), New Zealand (1973), Poland (1996), and Slovak Republic (2000). The OECD aims to enlarge its membership and to find the means to integrate nonmember governments

into its organization. The secretary-general Angel Gurria said he envisioned the OECD as the "secretariat of the process of globalization."

The basic document of OECD is the Convention of December 1960. It states that the organization shall promote policies (1) to achieve the highest sustainable economic growth and employment and a rising standard of living while maintaining financial stability, (2) to contribute to sound economic expansion, and (3) to expand world trade on a multilateral, nondiscriminatory basis.

Members are obliged to promote the efficient use of their economic, scientific, and technological resources; to pursue policies designed to achieve economic growth without thereby endangering the economies of other countries; to pursue efforts to reduce or abolish obstacles to the international movement of goods, services, payments, and capital; and to contribute to the economic development of member and nonmember countries by appropriate means. In addition, members must furnish the OECD with the information necessary for its tasks, consult together on a continuing basis, cooperate closely, and take coordinated action where appropriate.

The OECD may make decisions that are binding on all members. In addition, it may make recommendations to members and enter into agreements with member and nonmember states. Binding decisions and recommendations have to be made by mutual agreement. Abstention from voting does not invalidate the decision or recommendation, which is applicable to the other members but not to the abstaining member. This consensus method is the standard decision procedure of international organizations.

The OECD is entitled to diplomatic privileges and immunities, as are officials and member country representatives. The OECD is exempt from paying taxes related to its operating revenues and expenses.

Organization, Output, and Analytical Perspectives The supreme body of OECD is the Council, which is composed of all members and one representative of the European Union (EU) commission. It is a permanent intergovernmental conference that meets in sessions of ministers or permanent representatives. Representatives of the members meet in committees to discuss and review policies in various fields. There are about 200 committees, working groups, and experts groups. Each year some 40,000 senior officials from national administrations attend OECD committee meetings, where they request or review work of the secretariat or respond to requests by the secretariat. The OECD budget is financed by the member states. Their contributions are proportional to the size of their economies, with a floor of 0.1 percent and a cap of about 25 percent of total contributions. In 2006, the contributions of the United States, Japan, and Germany made up respectively 25, 17.5, and 9.4 percent of total contributions, while Iceland, Luxembourg, and the Slovak Republic each paid less than 0.25 percent of all contributions. Contributions make up about 80 percent of all operating revenues, which amounted to about U.S. $500 million per annum in 2005.

The secretary-general heads the secretariat. The Council elects him or her for a term of five years. He must neither seek nor receive instructions from any national government. The OECD has had five secretary-generals since its establishment: Thorkil Kristensen of Denmark (1961 69), Emil van Lennep of the Netherlands (1969 84), Jean-Claude Paye from France (1984 96), Donald Johnston of Canada (1996 2006), and Angel Gurria of Mexico (2006). Each secretary-general has had academic training in economics or law, a career in public service as a member of government or senior civil servant, and a political leaning toward economic liberalism.

The OECD has a staff of about 2,000, including about 700 economists, lawyers, and other professionals. The secretariat is structured by departments or directorates, working in parallel to the committees they serve.

Unlike the World Bank or the International Monetary Fund, the OECD does not dispense money to national governments. Instead, it uses its resources for gathering and analyzing data, as well as for publications and conferences. Major products of OECD are statistical data of great quality, analyses of

such data, the development of standards on matters such as corporate governance, and policy reviews. Among the most prominent of these reviews are the economic surveys. An economic survey is published every 1.5 to 2 years for each OECD member and for some large nonmember countries. The surveys analyze national economic policy, discuss policy options, and make recommendations.

The making of an economic survey reflects the political processes and structures within the OECD. The economics department, probably the most important of the 12 departments of the secretariat, organizes the procedure that leads to an economic survey. Within the economics department the Country Studies Branch, with its 15 country desks, is responsible for preparing the review of economic policy in close discussion with senior servants from the country under review. The draft of the economic survey is then submitted to the Economic and Development Review Committee (EDRC), composed of representatives of member countries. Two countries are responsible for the examination. They ask critical questions addressed to the delegation of the country under review. After the interaction of national civil servants, country desks, and examining countries, and after examination in front of the EDRC, the draft is revised and the members of the committee may comment on it. Discussion continues until the committee of representatives of 30 national governments makes a unanimous decision about the publication of the review.

As this example shows, continuous discussion, peer review, and consensual decision making are at the heart of this process, which requires representatives of national governments with very different political leanings to agree on analyses and recommendations for the economic policy of member countries.

The analytical perspectives of OECD have changed over time. When the organization started out it was concerned with core economic variables and had many intellectual affinities to Keynesian thinking. Beginning in the late 1970s it took a more monetarist outlook, and since the 1990s the OECD has broadened its former economic focus to include social and education policies and their different national configurations.

OECD and National Economic and Social Policy One of the major tools that the OECD uses to affect national economic and social policies is its data sets and reports produced by the secretariat in Paris. These reports point to shortcomings in national policy and describe alternative policies that are effective in other countries. The organization also diffuses ideas about good economic governance within an epistemic community of economists, lawyers, and social scientists from the OECD Secretariat and from national governments.

Finally, although the OECD issues standards, recommendations, and decisions, it does not employ sanctions if member countries ignore or violate these norms. "Naming and shaming" of such countries sometimes is taken up by national mass media and politicians. Analyses of the impact of the economic surveys have shown, however, that this is a rare event, especially for large countries. The OECD method of pressure by discussion, peer review, consensual decision taking, and naming and shaming has become a model for other organizations such as the EU, which adopted the OECD method as its "Open Method of Coordination" (OMC) as an alternative to less efficient and more problematic hierarchical methods of political steering. The EU applied OMC for the first time in 1997 and officially endorsed it as an EU method in 2000.

See also aid, international

FURTHER READING

Armingeon, Klaus, and Michelle Beyeler, eds. 2004. *The OECD and European Welfare States*. Cheltenham, UK: Edward Elgar. This volume contains 13 comparable country studies of the OECD's impacts on national welfare policy. An additional chapter describes the in ternal processes within the OECD, and a further com parative chapter draws conclusions based on the data of the country analyses.

Deacon, Bob. 2007. *Global Social Policy and Governance*. Los Angeles: Sage Publications. This book describes institutions of global social policies such as the World Bank, the United Nations, the International Monetary

Fund, and international nonstate actors. One section deals with the OECD.

Pagani, Fabrizio. 2002. *Peer Review: A Tool for Cooperation and Change. An Analysis of an OECD Working Method.* Paris: OECD. This paper describes the peer review process, which is the standard working method of the OECD when it evaluates countries and policies.

Sullivan, Scott. 2002. *From War to Wealth: 50 Years of In novation.* Paris: OECD. This book is an official "birth day book" on the occasion of the OECD's 50th anni versary.

KLAUS ARMINGEON

■ Organization of the Petroleum Exporting Countries (OPEC)

The Organization of the Petroleum Exporting Countries (OPEC) is an international cartel of oil-producing states that has attempted, with varied success, to manipulate world oil prices. OPEC was founded in 1960 by Iran, Iraq, Kuwait, Saudi Arabia, and Venezuela, a group of major oil-producing countries that wished to coordinate national petroleum policies and forge a more united front in dealings with the multinational oil companies that were licensed to produce and export petroleum from their lands. Within the next dozen years, eight additional members (Algeria, Ecuador, Gabon, Indonesia, Libya, Nigeria, Qatar, and the United Arab Emirates) joined in, which brought the total membership of OPEC to 13 by 1973. At that time, the combined membership of OPEC accounted for more than half of worldwide crude oil production. Two small producers (Gabon and Ecuador) withdrew during the 1990s and Angola joined in 2007, bringing OPEC membership to its current level of 12 countries.

As with any cartel, OPEC's ability to hold the price of oil above the competitive level is dependent on barriers to entry, which in this case hinge on OPEC's dominant ownership and control of low-cost oil reserves. By accident of nature, some 75 percent of the world's proved reserves of crude oil are located in OPEC nations. Proved reserves constitute that portion of the ultimate resource base that has already been discovered and is commercially producible. Additional reserves can and will be developed through exploration, discovery, and development of new fields, but this process has become increasingly difficult and expensive even more so outside the OPEC nations than within. Thus, while production of crude oil from non-OPEC sources does expand in response to the higher prices that result when cartel members restrict output, the scope for this is limited and will remain so. Moreover, OPEC's coordinated efforts to manipulate the price of oil are protected from antitrust enforcement and legal intervention by the sovereign rights of its members.

Economists have debated and tested various theories about how OPEC actually goes about exerting its influence on the market, whether through the independent initiatives of individual members, actions and strategies undertaken by semiautonomous coalitions working within the larger organization, or concerted plans embraced and executed by the organization as a whole. Some researchers might question whether OPEC has ever managed to operate successfully in the manner of a classic cartel. Whatever one's opinions on those matters, OPEC members have restricted production in ways that are unrelated to the inherent scarcity of crude oil. Although OPEC's proved oil reserves rose steadily between 1973 and 1985, production was cut by nearly half during that 12-year interval, falling from 31 million barrels per day (mbpd) in 1973 to an all-time low of 16 mbpd in 1985. As of late 2007, OPEC continued to hold production below the 1973 level, although the proved oil reserves of OPEC members had doubled in volume since then and total worldwide consumption of crude oil had grown by roughly 50 percent.

Evolution of OPEC The history and development of OPEC can be viewed in three phases. During the first phase (1960 70), OPEC's primary objective was to win for its members a larger *share* of the oil profits that private companies generated within their territory. The stated goal of increasing government take from 50 percent to 80 percent of total profits was pursued largely through the imposition of tax and administrative reforms by individual OPEC

members, including the introduction of fictional "tax reference prices" that boosted the tax base, and therefore government take, without altering the stated tax rate and without much impact on the market price of oil. During this phase, there was no direct attempt by OPEC to raise the overall level of world oil prices, and perhaps there was not even the realization that such a feat would be possible. In those early years, OPEC was concerned with winning for itself a bigger share of the pie, rather than growing the size of the pie.

The second phase (1970 82) saw greater reliance on collective deliberations and coordinated actions designed to reverse a long period of decline in world oil prices (and therefore tax revenues) that had set in after World War II. These efforts began with a series of dictated agreements (the so-called Teheran-Tripoli agreements of 1970 71) by which the OPEC members unilaterally raised posted tax reference prices by 21 percent. The members also announced that further increases could and would be imposed as they saw fit under the doctrine of "changing circumstances," one of which was the declining exchange value of the dollar, the currency in which oil prices were denominated. Indeed, it was during a special OPEC conference convened to review these matters that the October 1973 Arab-Israeli war broke out, which prompted the Arab members of OPEC to declare an embargo on sales to Israel's allies (the United States and the Netherlands). Although the embargo did not have much effect on actual deliveries of oil to those countries, and was soon rescinded, this bold move panicked the markets and fueled a speculative demand for oil inventories, which ultimately drove prices in the spot market to unprecedented levels and taught OPEC ministers something about the value of their oil.

By 1974, the "official" OPEC price had reached $11.25 per barrel, a startling increase from the $2.18 price level that had been established just two years before. By 1975, the posted price was no longer merely a fictional "tax reference" price used by OPEC members to compute their share of company profits. Indeed, the multinational companies were mostly removed from the equation by a wave of nationalizations that began in earnest in 1974, after which OPEC members sold their oil outright to whichever customers were willing to pay the official price. The posted price was successively increased during the 1970s by collective agreement of the OPEC ministers, but the real price of oil actually declined as the decade progressed since the posted price failed to keep pace with accelerating inflation.

Such was the state of affairs at the onset of the Iranian Revolution, when the expulsion of foreign oil field service firms and a series of labor strikes in 1978 and 1979 disrupted Iranian output. Disruptions spread to Iraq in 1980 with the outbreak of the Iran-Iraq war. Again the market panicked, and again the OPEC members were taught something about the value of their oil. By October 1981, the posted price of OPEC oil reached $34 per barrel (which in real terms still represents the all-time high).

A sharp downturn in the oil market led to the third (and current) phase in OPEC's evolution. Already by 1982, individual OPEC members were offering customers large discounts below the "official" OPEC price in order to maintain or even increase their share of what had become a dwindling market. Sluggish OPEC sales and falling prices were the product of reduced consumption and rising non-OPEC oil supplies, both spurred by the price shocks of the 1970s. To deal with the growing surplus of oil in the marketplace, OPEC adopted in March 1983 a formal system of production allocations that imposed for the first time individual ceilings on the output of each member. During this third phase of OPEC's development, OPEC members met at regular intervals (and sometimes more frequently on an emergency basis) to review market conditions and adjust members' quotas as needed to support or "defend" the market price within a desired range. This phase of OPEC's history is the one that most resembles the textbook example of a cartel, at least outwardly.

OPEC's Future Prospects When judging OPEC's past success or contemplating its future course of action, several things must be kept in mind. Foremost is the fact that any system of output restraint is vulnerable to the classic free-rider problem.

OPEC as a whole may benefit from reducing total output, but each member has an incentive to produce beyond its assigned quota. From the individual member's point of view, marginal revenue from incremental sales exceeds the marginal cost of extraction, which creates the temptation to cheat. Cartel membership is most beneficial to a producer when other members are doing the hard work. But if they will not, who will? Without a system to detect and punish cheating, the cartel is hampered by a prisoner's dilemma in which the dominant strategy for most, if not all, members is to ignore their assigned production quotas.

In fact, OPEC lacks an effective means to monitor, detect, and punish members who exceed their quotas. A monthly chart of OPEC's combined crude oil production level relative to the agreed ceiling indicates the scope and persistence of this problem (see figure 1). Compliance has been sporadic. Since the inception of the quota system, total OPEC production of crude oil has exceeded the ceiling by 4 percent on average, but on numerous occasions the excess has run to 15 percent or more. For the most part, compliance has been achieved only during episodes (like 2005 6) when the production ceiling itself pushed the limits of each member's available production capacity.

A second factor that confounds OPEC's attempt to manage the market price is the lack of timely and accurate information about changes in the level of demand for oil and the availability of non-OPEC oil supplies. Several forecasts of demand and supply are available at any given time (including those prepared by the U.S. Energy Information Administration, the International Energy Agency, and by the OPEC Secretariat itself), but the precision of these forecasts is low and surprises are frequent. For example, none anticipated the surge in Asian demand that triggered the sudden tightening of oil markets in 2005. OPEC's forecasting problem is compounded by the fact that several years may elapse, due to rigidities in both supply and demand, before the full impact of a price change can be observed so if a mistake is made, it may go undetected for several years and then take several years more to rectify.

Even if perfect information about future market conditions were available, there is no assurance that the interests of individual OPEC members could be easily aligned around a single "correct" price or production target. OPEC has very limited means by which to redistribute earnings among members. Therefore, any given set of quotas determines not only the overall profit of OPEC, but also the individual revenues that accrue to each member.

If the members were more homogeneous demographically and economically, the problem of misaligned interests would be less severe. As things happen, however, large volumes of low-cost reserves are concentrated in certain countries with small populations and relatively high incomes (e.g., Kuwait, Saudi Arabia, and the United Arab Emirates), while smaller volumes of higher-cost reserves are found in populous and relatively poor countries (e.g., Nigeria, Indonesia, and Venezuela). Table 1 sets forth some of the more salient differences among the members of OPEC. The potential for conflicting interests involves not only the question of which members "deserve" larger quotas, but what is the preferred market price level for OPEC oil.

What price would the respective members of the cartel like to see? Members with low-cost, long-lived reserves will take a long view of the future and may be reluctant to push prices too high given the fear of induced technological innovations that would usher
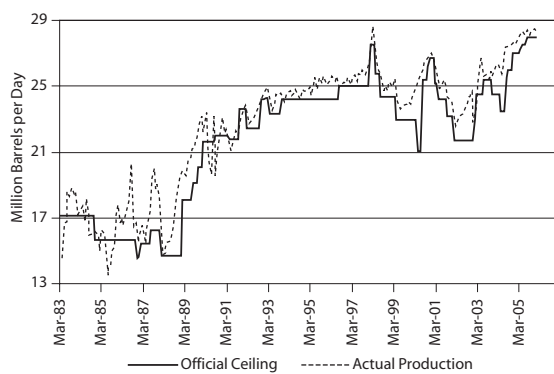
Figure 1

OPEC production compliance (monthly data). *Source:* ceilings, OPEC annual statistical bulletins; actuals, U.S. Energy Information Administration.

Table 1
Differences among OPEC members

Member	since	GDP ($ per capita)	Value of oil exports ($ per capita)	Proved oil reserves (bbl per capita)	Crude oil production (bbl per capita)	Reserves to production ratio (years)
Algeria	1969	3,113	999	373	15	25
Indonesia	1962	1,290	42	20	2	11
Iran*	1960	2,863	704	1,986	22	91
Iraq*	1960	1,063	812	3,989	24	165
Kuwait*	1960	27,028	15,429	36,775	340	108
Libya	1962	6,618	4,839	7,084	106	67
Nigeria	1971	752	355	275	7	42
Qatar	1961	45,937	22,614	18,455	339	54
Saudi Arabia*	1960	12,931	6,876	11,029	143	77
UAE	1967	29,367	11,044	21,733	193	113
Venezuela*	1960	5,240	1,796	2,990	43	70
OPEC average		2,649	941	1,660	21	81

Source: OPEC Annual Statistical Bulletin, 2005
Note: Angola joined OPEC in 2007; data for that country are not included here.
*Founding member of OPEC

in new forms of energy (or energy conservation) that eventually may compete against OPEC. Members holding fewer reserves and shorter horizons are less vulnerable to this type of risk and therefore perhaps less averse to high prices. Internal divisions between "price hawks" and "price doves" have been observed previously and will likely surface within OPEC again.

A final factor that looms large in the future of OPEC is the role to be played by serendipitous events and geopolitical tensions. A large portion of OPEC's apparent historical impact on the price of oil has come about not as the result of deliberate plans crafted by a purposeful cartel, but as the by-product of clashing national agendas that encompass far more than the petroleum sector. Since the early 1970s, most of the idle capacity held by OPEC members has been involuntary   taken out of production due to military conflict. Much of the hard work that any cartel has to do   commanding the determination and discipline to restrict output   has in OPEC's

case been provided fortuitously. For that reason, the ultimate strength and cohesion of OPEC had perhaps not yet been tested by the end of 2007.

The value of crude oil produced and sold on the world market exceeds $1 billion each day. Even a relatively small impact on the unit price of oil represents an enormous transfer of wealth between consumers and producers. Moreover, the disruptive impact of sudden price "shocks" and heightened volatility threatens the goal of sustained and steady global economic growth. As consumers, investors, and government officials continue to wrestle with these problems, it is no exaggeration to say that OPEC has left an indelible imprint on the world economy through its impact on the price of oil.

See also Gulf Cooperation Council; petroleum; primary products trade

FURTHER READING

Adelman, M. A. 1995. *The Genie Out of the Bottle: World Oil since 1970*. Cambridge, MA: MIT Press. A compre

hensive study of the world oil market and OPEC's central role within it.

. 2002. "World Oil Production and Prices: 1947 2000." *Quarterly Review of Economics and Finance* 42 (2): 169 91. A briefer and more recent examination of OPEC and its antecedents.

Amuzegar, Jahangir. 1999. *Managing the Oil Wealth: OPEC's Windfalls and Pitfalls.* London: I. B. Taurus. Insightful commentary on the financial challenges that confront oil dependent economies.

Claes, Dag Harald. 2001. *The Politics of Oil Producer Co operation.* Boulder, CO: Westview Press. A look at the basis for economic cooperation among oil producers from a political science perspective.

Crémer, Jacques, and Djavad Salehi Isfahani. 1991. *Models of the Oil Market.* Chur, Switzerland: Harwood Aca demic. Quantitative modeling of OPEC behavior.

Jones, Donald W., Paul Leiby, and Inja K. Paik. 2004. "Oil Price Shocks and the Macroeconomy: What Has Been Learned Since 1996." *Energy Journal* 25 (2): 1 32. A survey that also reveals much about what had been learned prior to 1996.

Parra, Francisco. 2004. *Oil Politics: A History of Modern Petroleum.* London: I. B. Taurus. A fascinating insider's guide to OPEC policies and politics.

Smith, James L. 2005. "Inscrutable OPEC? Behavioral Tests of the Cartel Hypothesis." *Energy Journal* 26 (1): 51 82. An interpretation of OPEC production patterns in terms of competitive versus collusive behavior.

JAMES L. SMITH

■ original sin

The term *original sin*, obviously an allusion to Christian theology, was coined by economists Barry Eichengreen and Ricardo Hausmann to refer to the inability of a particular country (most notably, emerging-market economies) to borrow funds in international capital markets in its own currency. The inability to borrow in domestic currency units necessarily forces the country in question to raise external funds using a foreign convertible currency instead, such as the U.S. dollar, the British pound, the euro, or the Japanese yen. Countries that choose to borrow external funds in this fashion expose themselves to a currency mismatch problem in their balance sheets, as their liabilities are denominated in foreign currency units, while their assets are de-nominated in domestic currency units. If there were no exchange rate movement over time, there would be no balance sheet valuation effects. If the domestic currency were to depreciate or be devalued, however, the country in question would experience a reduction in the market value of its assets, with little or no change in the value of its liabilities. As a result, the likelihood of insolvency and default increases. The risk of currency mismatch affects all sectors exposed to the problem: governments, banks, and even cor-porations.

Economic Impact of Original Sin There is a large and growing theoretical and empirical literature that analyzes the effect of original sin on macroeconomic performance. To be sure, standard models of the open economy do not directly address the issue of original sin. In these models, access to external debt (whether foreign-currency or domestic-currency denominated) is seen as beneficial, as this facilitates the country's ability to smooth out fluctuations in aggregate demand. A country that lacks the ability to borrow funds from the rest of the world is presum-ably worse off than it would otherwise be as it must depend solely on its own savings for financing con-sumption and investment expenditures. If domestic savings is insufficient for funding consumption and investment, access to external capital markets can help to finance the gap, thereby helping to maintain macroeconomic stability and growth. More recent models, however, show that when a large portion of the external debt is foreign-currency denominated, the likelihood of experiencing a balance of payments crisis may increase (e.g., Jeanne and Zettelmeyer 2005). Macroeconomic stability can therefore be seriously compromised when a country is exposed to original sin. In addition, monetary and fiscal policy tools can also be adversely affected by original sin. In particular, original sin could make monetary policy more complex, causing it to lose some of its effec-tiveness in influencing output (Cespedes, Chang, and Velasco 2005; Rajan 2006).

Empirically, many researchers find that original sin actually amplifies macroeconomic volatility. In particular, the higher the level of original sin, the more likely is the country to suffer from instability in output and capital flows. Many researchers see original sin as at least partly responsible for many banking crises in recent decades (Kaminsky and Reinhart 1999). The reasoning behind this belief is that the banking sector in many emerging market economies is often exposed to currency fluctuations: deposits are often denominated in a foreign, convertible currency (in order to attract funds that would otherwise go abroad), while bank assets are primarily loans denominated in domestic currency. When the currency depreciates or is devalued, the market value of bank loans declines; the value of liabilities, on the other hand, remains roughly unchanged, thereby compromising the solvency of the banking sector. The effect of such an event on the rest of the economy can be quite devastating as banks, in an effort to increase liquidity, contract their lending, thereby creating a "credit crunch" that amplifies an economic downturn. Indeed, many researchers argue that original sin is largely responsible for some of the most devastating financial crises in recent history, such as the Asian financial crisis of 1997–98.

Reasons behind Original Sin Given that original sin appears to cause a significant amount of macroeconomic problems, it is natural to ask why countries borrow in foreign currency units. The literature mentions several reasons, including (1) weak economic credibility, (2) weak financial institutions, and (3) structural impediments to domestic currency debt issuance. First, emerging market economies typically lack a strong track record of economic performance. For example, emerging market economies tend to have long-lasting inflation problems. Because of lack of monetary credibility, foreign investors may be reluctant to lend to such countries in domestic currency units, demanding instead that countries issue foreign-currency denominated debt (Jeanne 2005). Second, it is likely that foreign investors are uneasy about holding the debt of countries with weak institutions, such as a poor track record on the enforcement of contracts, and poorly enforced property rights. Facing this constraint, countries may choose to issue debt denominated in foreign currency units as a way to signal to potential investors that they are committed to keep their end of the obligation (Reinhart, Rogoff, and Savastano 2003). Third, the international financial architecture simply does not allow many countries to issue debt in domestic currency units. In particular, some researchers have noted that countries that borrow in international markets using their own currency tend to have large economies and deep financial systems (Flandreau and Sussman 2005). If so, this would imply that issues related to economies of scale, network externalities, or liquidity allow only a handful of currencies to be effectively used in international financial contracts.

See also balance sheet approach/effects; banking crisis; conflicted virtue; convertibility; currency crisis; currency substitution and dollarization; dollar standard; dominant currency; financial crisis; reserve currency

FURTHER READING

Calvo, G., and C. Reinhart. 2002. "Fear of Floating." *Quarterly Journal of Economics* 117: 379–408. Investigates whether countries that claim to have a floating exchange rate actually allow their currencies to fluctuate. They find that, for the most part, countries do not. They argue that the "fear of floating" appears to be epidemic.

Cespedes, L. F., R. Chang, and A. Velasco. 2005. "Must Original Sin Cause Macroeconomic Damnation?" In *Other People's Money: Debt Denomination and Financial Instability in Emerging Market Economies*, edited by B. Eichengreen and R. Hausmann. Chicago: University of Chicago Press, 48–67. Presents a theoretical model showing that original sin can undermine the effectiveness of fiscal and monetary policy tools, thereby compromising a country's ability to stabilize output.

Eichengreen, B., and R. Hausmann. 1999. "Exchange Rates and Financial Fragility." NBER Working Paper No. 7418. Cambridge, MA: National Bureau of Economic Research. Introduces for the first time the term *original sin* as one of the avenues through which exchange rate volatility can induce financial fragility.

———. 2005. "Introduction: Debt Denomination and Financial Instability in Emerging Market Economies."

In *Other People's Money: Debt Denomination and Financial Instability in Emerging Market Economies*, edited by B. Eichengreen and R. Hausmann. Chicago: University of Chicago Press, 3–12. Offers an excellent exposition on the issues pertaining to original sin. The introductory chapter summarizes the main contributions of the papers included in the book.

Eichengreen, B., R. Hausmann, and U. Panizza. 2005. "The Pain of Original Sin." In *Other People's Money: Debt Denomination and Financial Instability in Emerging Market Economies*, edited by B. Eichengreen and R. Hausmann. Chicago: University of Chicago Press, 13–47. Presents empirical evidence showing that financial and macroeconomic instability in emerging market economies tends to increase with original sin problems.

Flandreau, M., and N. Sussman. 2005. "Old Sins: Exchange Clauses and European Foreign Lending in the Nineteenth Century." In *Other People's Money: Debt Denomination and Financial Instability in Emerging Market Economies*, edited by B. Eichengreen and R. Hausmann. Chicago: University of Chicago Press, 154–89. Argues that the main cause of original sin is not the lack of credibility and discipline of governments, as others have argued, but rather it is the result of historical path dependence and market liquidity.

Jeanne, O. 2005. "Why Do Emerging Economies Borrow in Foreign Currency?" In *Other People's Money: Debt Denomination and Financial Instability in Emerging Market Economies*, edited by B. Eichengreen and R. Hausmann. Chicago: University of Chicago Press, 190–217. Explores the causes of original sin and argues that original sin in private debt markets is mainly the result of monetary policy.

Jeanne, O., and J. Zettelmeyer. 2005. "Original Sin, Balance Sheet Crises, and the Roles of International Lending." In *Other People's Money: Debt Denomination and Financial Instability in Emerging Market Economies*, edited by B. Eichengreen and R. Hausmann. Chicago: University of Chicago Press, 95–121. Explores two classes of models and finds that in both of them crises can be self fulfilling prophesies. The paper also investigates the role of international lending in ameliorating the effect of financial crises stemming from original sin problems.

Kaminsky, G., and C. Reinhart. 1999. "The Twin Crises: The Causes of Banking and Balance of Payments Problems." *American Economic Review* 89 (3): 473–500. Investigates the links between banking crises and currency crises. Its main finding is that banking problems typically precede currency crises, and once a currency crisis is triggered, it amplifies banking problems, culminating in a banking crisis.

Rajan, R. S. 2006. "Managing New Style Currency Crisis: The Swan Diagram Approach Revisited." *Journal of International Development* 18: 1–24. Presents a model useful for understanding how countries can manage economic disruptions stemming from crises driven by original sin problems.

Reinhart, C., K. Rogoff, and M. Savastano. 2003. "Debt Intolerance." *Brookings Papers on Economic Activity* 1: 1–74. Quantifies a country's vulnerability to a debt crisis as it accumulates external debt. Its main finding is that having experienced debt problems in the past makes a country more susceptible to external debt crises in the future.

CARLOS D. RAMIREZ

■ outsourcing/offshoring

The terms *foreign outsourcing* or *offshoring* apply when the components of a good or service are produced in several countries. The term *offshoring* often refers to a company moving some of its operations overseas, but retaining ownership of those operations. Intel, for example, produces microchips in China and Costa Rica using subsidiaries that it owns, so these production activities have moved offshore. In contrast, *outsourcing* refers to moving activities outside of a firm (which could be to another firm in the same country, as with *domestic* outsourcing, or to another firm in another country, as with *foreign* outsourcing). Mattel, for example, arranges for the production of the Barbie doll in several different countries, so it is engaged in foreign outsourcing. Unlike Intel, however, Mattel does not actually own the firms in those countries. In this entry we will not be concerned with the distinction between foreign outsourcing or offshoring, and use either term to refer to shifting activities to another country.

While there are historical examples of companies doing some of their production in another country, outsourcing is generally thought to be a feature of the modern world economy made possible by improvements in international trade, transportation, and communication. Indeed, the earliest known use of the word *outsourcing* in a published source is from an American auto executive in the *Journal of the Royal Society of Arts*, 1979, who wrote: "We are so short of professional engineers in the motor industry that we are having to outsource design work to Germany" (Safire 2004). This example shows that outsourcing may involve the shifting of service activities (like design work) overseas, in addition to the shifting of production activities (like making the Barbie doll) overseas. In this entry we first concentrate on the shifting of *production activities* to other countries and then discuss *service outsourcing*.

Measures of Outsourcing There are several approaches that can be used to measure the amount of outsourcing. One approach is to look at "processing trade," which is defined by customs offices as the import of intermediate inputs for processing and subsequent reexport of the final product. This activity has grown enormously in China, for which Hong Kong often serves as an intermediary. For example, between 1988 and 1998, processing exports grew from $12.4 billion to $97.2 billion, or from about one-third to over one-half of total Chinese exports (Feenstra and Hanson 2004). This outward processing serves newly industrialized countries in Asia, but also developed countries such as the United States, Japan, and countries in Europe. Between the industrialized countries, too, there has been an increase in processing trade. Görg (2000) reports on the increase in U.S. processing trade with the European Union between 1988 and 1994. He finds that U.S. processing imports into these countries (as a share of their total U.S. imports) increased slightly from 17.7 percent to 19.8 percent, but this same ratio increased more significantly from 13.7 percent to 23.7 percent for U.S. exports into the "periphery" countries of Greece, Ireland, Portugal and Spain.

Another way to measure foreign outsourcing is by the amount of imported intermediate inputs, which can be estimated by using the purchases of each type of input and multiplying it by the economywide import share for that input. Adding overall inputs used within each industry, we obtain estimated imported inputs, which can then be expressed relative to total intermediate input purchases. Feenstra and Hanson (1999) perform this calculation for U.S. manufacturing industries and find that imported inputs increased from 6.5 percent of total intermediate purchases in 1972 to 8.5 percent in 1979, and 11.6 percent in 1990. Campa and Goldberg (1997) make the same calculation for Canada, Japan, the United Kingdom, and the United States. The United States shows a doubling of the share of imported inputs between 1975 and 1995 for all manufacturing, from 4.1 percent to 8.2 percent, though it is still at a low level compared with other countries. Canada shows an increase in the share of intermediate inputs from 15.9 percent to 20.2 percent from 1974 to 1993, and the United Kingdom shows an especially large increase in this share, rising from 13.4 percent to 21.6 percent over the same years. The exception is Japan, where the share of imported inputs in manufacturing fell. With that single exception, the increased use of imported inputs was a characteristic feature of many industrial countries during the 1980s and 1990s.

Effect of Outsourcing on Wages: Evidence from the 1980s Much of the academic and policy interest in outsourcing is due to its potential effect on wages and employment. During the 1980s there was a surprising movement in wages and employment in the United States and other countries. During that decade, the real wage of less-skilled workers (with high-school education or less) fell in the United States, whereas the real wages of the more highly skilled workers (college graduates) rose. Therefore, the ratio of the skilled wage divided by the unskilled wage or the relative wage of skilled/unskilled workers rose. At the same time, the relative employment of skilled/unskilled workers also went up, especially in manufacturing. That pattern is surprising because normally when the relative wage of skilled/unskilled workers rises, we expect that companies will hire fewer skilled employees (since they are more ex-

pensive); instead, the opposite happened and companies hired relatively more skilled workers. The only explanation for this pattern is that the relative *demand* for skilled workers must have increased, especially in the manufacturing sector. What factors can explain this increase in the relative demand for skilled workers?

Two factors that can explain the increase in relative demand for skilled workers are: (1) the increased use of computers and other high-technology equipment, and thus an increase in the skilled workers needed to operate them; and (2) outsourcing. To understand how outsourcing will increase the relative demand for skilled labor, we use the "value chain" of a firm, which includes all the activities involved in the production of a good or service, from research and development (R&D) to assembly to marketing and after-sales service. For the purpose of modeling outsourcing, rather than arranging activities in the order they are actually performed, we instead arrange them in increasing ratio of skilled/unskilled labor used in each activity, as shown in figure 1.

Assembly uses the least amount of skilled labor relative to unskilled labor in figure 1, followed by component production, then marketing and sales, and finally R&D. A firm that is outsourcing to another country with lower relative wages for unskilled labor will want to send those activities using the most unskilled labor. So in figure 1, activities to the left of the line AA will be sent offshore to the foreign country, while activities to the right of the line AA will be performed at home.

Now suppose that the home firm wishes to offshore more activities. The reason for this could be a trade agreement with the foreign country, leading to reduced tariffs; or improvement in the infrastructure in the foreign country, leading to reduced costs there; or an increase in costs at home. When deciding what extra activities to offshore, the firm will look to those activities that were just on the borderline of being outsourced before, that is, those activities just to the right of the line AA, which used to be profitably performed at home but now are shifted abroad. The borderline between the activities performed at home and abroad therefore shifts from the line AA to the line BB.
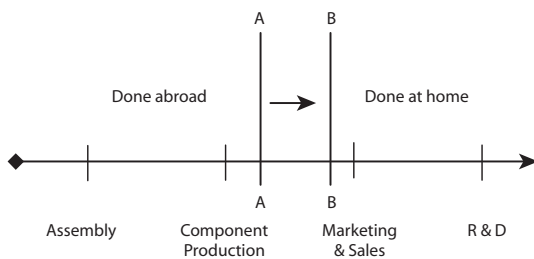
What is the impact of this increase in outsourcing on the relative demand for skilled labor at home and abroad? Notice that the activities no longer performed at home (i.e., those in between AA and BB) are *less* skill-intensive than the activities still done there (those to the right of BB). This means that the range of activities now done at home is more skilled-labor intensive, on average, than the set of activities formerly done at home. For this reason, the relative demand for skilled labor at home increases, as occurred in the United States during the 1980s. That increase in demand will also increase the relative wage for skilled labor.

What about in the foreign country? The activities that are newly sent offshore (those in between AA and BB) are *more* skill-intensive than the activities that were initially outsourced to the foreign country (those to the left of AA). That means that the range of activities now done abroad is more skilled-labor intensive, on average, than the set of activities formerly done there. For this reason, the relative demand for skilled labor in the foreign country also increases. With this increase in the relative demand for skilled labor, the relative wage of skilled labor also increases in the foreign country. That outcome occurred in Mexico, for example, during the 1980s, as well as in Hong Kong.

Outsourcing versus Technological Change By shifting activities from one country to the other, outsourcing can increase the relative demand for skilled labor in *both* countries, as has actually occurred in a number of industrial and developing countries. However, the same result can occur from skill-biased technological change, such as the

Figure 1

Outsourcing on the value chain

increased use of computers, which can increase the relative demand for skilled labor across countries. Given that outsourcing and skill-biased technological change both predict an increase in the relative wage of skilled labor, it becomes an empirical issue as to which is more important.

A study for the United States (Feenstra and Hanson 1999) seeks to explain the increase in the share of total wage payments going to nonproduction (skilled) versus production (unskilled) labor in U.S. manufacturing industries over the period 1979 90 and analyzes the increase in the relative wage of nonproduction labor over the same period. The study considers two possible explanations for the change in wages: outsourcing and the use of high-tech equipment such as computers. High-technology equipment can itself be measured in two ways: either as a fraction of the total capital equipment installed in each industry or as a fraction of new investment in capital that is devoted to computers and other high-tech devices.

Using the first measure of high-tech equipment (i.e., fraction of the capital stock), 20 23 percent of the increase in the share of wage payments going to nonproduction workers was explained by outsourcing, and 8 12 percent of the increase was explained by the growing use of high-tech capital. Thus, using the first measure of high-tech equipment, it appears that outsourcing was more important than high-tech capital in explaining the change in relative demand for skilled workers. The story is different, however, when the second measure of high-tech equipment (i.e., fraction of new investment) is used. In that case, outsourcing explains only 13 percent of the increase in the nonproduction share of wages, whereas high-tech investment explains 37 percent of that increase. So we see from these results that both outsourcing and high-tech equipment are important explanations for the increase in the relative wage of skilled labor in the United States, but which one is more important depends on how we measure the high-tech equipment.

Moving on to the increase in the relative wage of nonproduction workers, using the first measure of high-tech equipment (fraction of the capital stock),

21 27 percent of the increase in the relative wage of nonproduction workers was explained by outsourcing, and 29 32 percent of the increase was explained by the growing use of high-tech capital. Using the other measure of high-tech equipment (fraction of new investment), the large spending on high-tech equipment in new investment can explain *nearly all* (99 percent) of the increase in the relative wage for nonproduction workers, leaving little room for outsourcing to play much of a role (it explains only 12 percent of the increase in the relative wage). These results are lopsided enough that we might be skeptical of using new investment to measure high-tech equipment and therefore prefer the results using the capital stocks. Summing up, both outsourcing and high-tech equipment are important explanations for the increase in the relative wage of nonproduction/ production labor in U.S. manufacturing, but the relative contributions of the two measures are very sensitive to how we measure the high-tech equipment.

Trade Costs and Outsourcing across Firms The effects of outsourcing described in figure 1 can be thought of as occurring along the value chain of a firm. One natural way to examine these changes is by examining the impact of falling trade costs on manufacturing establishments with different characteristics. It is likely that as firms move production activities offshore, they will outsource the least skilled activities (as depicted in figure 1), or close the plants focusing on these activities.

Bernard, Jensen, and Schott (2006) examine the implications of falling trade costs on U.S. manufacturers, and specifically examine the channels by which trade affects the distribution of economic activity. They find when trade costs in an industry fall, plants are more likely to close. They also find that low productivity, nonexporting plants are more likely to die. This is one channel by which outsourcing can affect the distribution of economic activity. Falling trade costs tend to reduce the amount of economic activity at the low end of the productivity distribution. Because low-productivity plants also tend to be production-worker intensive, this change is likely to reduce the relative demand for unskilled workers.

Bernard, Jensen, and Schott (2006) also find that relatively high-productivity nonexporters in industries with falling trade costs are more likely to start exporting. The magnitude of the effect of falling trade costs on becoming an exporter is substantial. Because higher-productivity plants are more skilled-worker intensive, as these plants expand they will increase the relative demand for skilled workers. They also find that existing exporters increase their shipments abroad as trade costs fall. Because exporters have relatively high-productivity plants, the expansion of the high end of the productivity distribution will tend to raise aggregate productivity (even if no plants changes its productivity). Because exporters are skill- and capital-intensive, this will also tend to increase relative demand for these factor inputs. Finally, these authors find that plants in industries with falling trade costs have faster productivity growth, possibly due to increased outsourcing.

It should be noted that the productivity increase associated with outsourcing means that the real wage of workers (even the less-skilled workers) need not fall due to outsourcing. That result is shown in the model of Feenstra and Hanson (1996), from which figure 1 is drawn. The same result occurs more strongly in the recent model of Grossman and Rossi-Hansberg (2006), where the real wage of less-skilled worker are guaranteed to rise due to the productivity-enhancing effect of outsourcing.

Regional Variation in Wage Inequality in the United States The increase in wage inequality due to outsourcing has not been uniform across the United States. Bernard and Jensen (2000) find that the changing composition of employment in regional economies is strongly correlated with changes in wage inequality in the United States. Somewhat surprisingly, they find that while many regions in the United States experienced increasing wage inequality, some regions experienced *decreasing* wage inequality over the 1970–90 period.

Figure 2 shows the variation across states in changes in residual wage inequality for 1970–80 (upper map) and 1980–90 (lower map). States with large increases in wage inequality are in the traditional "Rust Belt," and states with decreases in wage
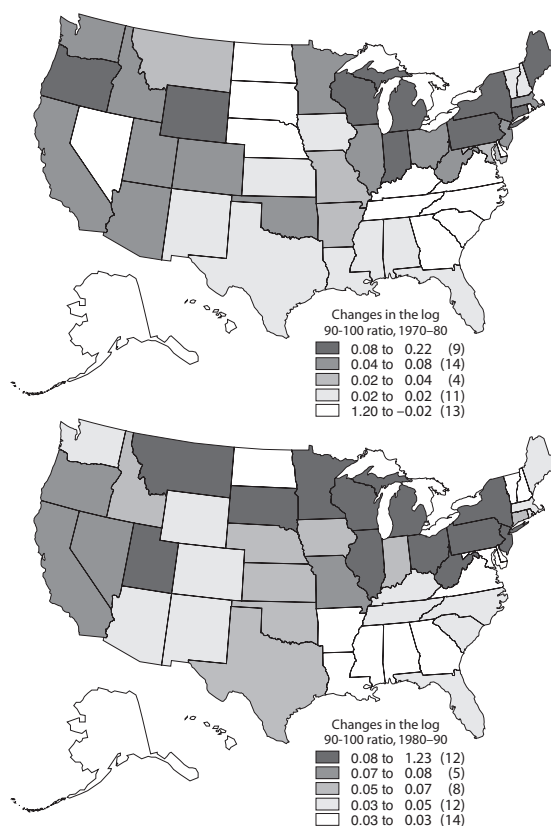
Figure 2

Changes in inequality by state. Source: Bernard and Jensen (2000).

inequality are in the Southeast. Bernard and Jensen find that decreases (increases) in the share of manufacturing sector employment in the durable goods sector is strongly correlated with increasing (decreasing) wage inequality.

One specific example of how these heterogeneous responses to trade pressures can affect regions differentially is the case of Appalachian manufacturing. Appalachian manufacturing is concentrated in low-wage, labor-intensive industries, and Appalachian manufacturers within those industries tend to be lower wage and lower productivity than plants in those industries elsewhere in the United States. This combination of industrial mix and plant production technology leaves the region particularly susceptible to import competition from low-wage countries. For

example, low-wage competition shows a more pronounced effect on Appalachian plants in terms of employment growth and plant failure than elsewhere in the United States. Plants in the Appalachian region have higher shutdown probabilities and lower employment growth when facing low-wage imports than do firms in the rest of the United States.

Service Outsourcing in Manufacturing: Evidence from the 1990s The patterns of wages and employment in U.S. manufacturing changed in the 1990s, with rising relative wages and employment for skilled workers. The relative wage of nonproduction/ production labor in U.S. manufacturing continued to increase from 1989 to 2000, but in addition, the relative employment of production workers *decreased*. A likely explanation for this new pattern of wage and employment is that it reflects service outsourcing from U.S. manufacturing. To the extent that the back-office jobs being outsourced from manufacturing use the lower-paid nonproduction workers, then the offshoring of those jobs could very well *raise the average* wage among nonproduction workers, while lowering their employment. So that pattern would be consistent with what has actually occurred in U.S. manufacturing. In the rest of our entry, then, we focus on service outsourcing.

Examples of service outsourcing from the manufacturing sector include the offshoring of services such as communication, finance, insurance, computer, and information services. Amiti and Wei (2006) report that in the United States, the amount of imported service inputs is small but growing. Measured as a share of total inputs purchased, imported services were 0.2 percent in 1992 (i.e., two-tenths of one percent of total inputs), and grew to 0.3 percent in 2000 (i.e., three-tenths of one percent). The fact that imported services are small does not necessarily prevent them from being important for productivity and employment.

In terms of the impact of service outsourcing and high-technology equipment on manufacturing productivity measured by value added per worker, over the eight years 1992 2000, Amiti and Wei find service outsourcing can explain about 11 percent of the total increase in productivity. Despite the small

amount of service imports, they find it explains a significant portion of productivity growth. It may be some of that productivity growth is actually due to domestic outsourcing of service activities, which is confounded in the data with foreign outsourcing of services. The contribution of service outsourcing can be compared to the offshoring of material inputs, which explains a further 5 percent of the total increase in productivity. Adding together these contributions, we see that these two factors explain about 16 percent of the increase in value added per worker, or as much as one-sixth of productivity growth. Since productivity rose by about 4 percent per year in manufacturing, these results show that outsourcing of services together with material inputs during the 1990s can explain two-thirds of a percentage point in productivity growth per year, which is economically important.

Offshoring's Impact on the Service Sector In the previous section we examined the impact of service outsourcing on the manufacturing sector. In this section we consider the potential impact of offshoring on the service sector. While work on trade in services is quite recent and hampered by less detailed information than trade in goods, Jensen and Kletzer (2006) provide evidence on the potential impact that trade in services will have. They use the geographic concentration of service activities within the United States to identify industries and occupations that appear traded across regions within the United States and classify these activities as tradable. They find that a significant share of total employment is in tradable service activities: in fact, more employment is in tradable professional and business services than in manufacturing. They also find that workers in tradable service activities are different than workers in nontradable activities, even in the same sector. Workers in tradable services have more education and significantly higher earnings (after controlling for observable differences). They examine recent trends in employment growth and find little evidence that tradable services have lower employment growth than no-tradable services, though they do find some evidence of higher displacement rates in tradable service activities.

These results suggest that the potential scope for tradable services is large enough to have a significant impact on the U.S. economy. It seems likely that increasing trade in services will have a similar effect as the increased trade in goods. Trade in services is likely to increase productivity through a number of channels, including closure of low-productivity service producers, entry into exporting of relatively high-productivity service providers, and expansion of service firms that already export. Similar to manufacturers, service exporters likely have superior operating characteristics and use a different skill mix than nonexporters. We expect increased trade in services will affect the relative demand for skilled and unskilled labor, though it may affect a different portion of the skill distribution than we witnessed in the manufacturing sector in the 1980s and 1990s.

See also fragmentation; internalization theory; trade and wages

FURTHER READING

Amiti, Mary, and Shang Jin Wei. 2006. "Service Off shoring and Productivity: Evidence from the United States." NBER Working Paper No. 11926. Cambridge, MA: National Bureau of Economic Research. Estimates the impact of service outsourcing from the manufac turing sector on wages and productivity in the United States.

Bernard, Andrew B., and J. Bradford Jensen. 2000. "Un derstanding Increasing and Decreasing Wage Inequal ity." In *The Impact of International Trade on Wages*, edited by Robert Feenstra. Chicago: University of Chi cago Press, 227 61. Examines the regional variation in wage inequality in the United States.

Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott. 2006. "Trade Costs, Firms, and Productivity." *Journal of Monetary Economics* 53 (5): 917 37. Ex amines the impact of falling trade costs on firm entry, exit, and productivity.

Campa, Jose, and Linda Goldberg. 1997. "The Evolving External Orientation of Manufacturing Industries: Evidence from Four Countries." NBER Working Paper No. 5919. Cambridge, MA: National Bureau of Eco nomic Research; *Economic Policy Review,* The Federal Reserve Bank of New York (July). Presents evidence on the amount of outsourcing for four industrial countries: Canada, Japan, the United Kingdom, and the United States.

Feenstra, Robert C., and Gordon H. Hanson. 1996. "For eign Investment, Outsourcing, and Relative Wages." In *The Political Economy of Trade Policy: Papers in Honor of Jagdish Bhagwati*, edited by R. C. Feenstra, G. M. Grossman, and D. A. Irwin. Cambridge, MA: MIT Press, 89 127. Presents a model of outsourcing and some preliminary empirical results.

. 1999. "The Impact of Outsourcing and High Technology Capital on Wages: Estimates for the U.S., 1979 1990." *Quarterly Journal of Economics* 114 (3) (August): 907 40. Estimates the impact of outsourcing and computer use on wages in U.S. manufacturing.

. 2004. "Intermediaries in Entrepôt Trade: Hong Kong Re Exports of Chinese Goods." *Journal of Eco nomics and Management Strategy* 13 (1) (spring): 3 35. Discusses Chinese processing trade and intermediation by Hong Kong.

Görg, Holger. 2000. "Fragmentation and Trade: U.S. In ward Processing Trade in the EU." *Weltwirtschaftliches Archiv* (*Review of World Economics*): 136: 403 22. Dis cusses processing trade between the United States and Europe.

Grossman, Gene, and Esteban Rossi Hansberg. 2006. "Trading Tasks: A Simple Model of Outsourcing." NBER Working Paper No. 12721. Cambridge, MA: National Bureau of Economic Research. A theoretical exposition of outsourcing in a two sector model.

Jensen, J. Bradford, and Lori Kletzer. 2006. "Tradable Services: Understanding the Scope and Impact of Ser vices Offshoring." In *Offshoring White Collar Work Issues and Implications*, edited by Lael Brainard and Susan M. Collins (Washington, DC: Brookings In stitute Trade Forum 2005), 75 134. Examines out sourcing from the service sector of the United States.

Safire, William. 2004. "On Language." *New York Times Magazine* (March 21), 30. Discusses the earliest known use of the word *outsourcing* in a published source.

ROBERT C. FEENSTRA AND J. BRADFORD JENSEN

◼ parallel imports

Parallel imports (also known as gray products) are products imported into a country without the authorization of the intellectual property right owner. They arise when international price differences exceed the costs of transporting the products across borders. No precise data exist on parallel imports, but their volume is considered to be large (up to a market share of 20 percent in some countries and for some products), especially for products such as musical recordings, cosmetics, fragrances, soft drinks, clothing, confectionery, automobiles, motorcycles, consumer electronics, and pharmaceuticals (Maskus 2000).

The international system of intellectual property rights allows each country to establish its own legal regime concerning parallel trade. It is based on the concept of exhaustion of property rights (known as "first-sale" doctrine in the United States), whereby countries choose the geographical area (national, regional, international) within which the owners of intellectual property rights lose (and thus "exhaust") their rights on products after first sales. Most countries have a regime of *national* exhaustion of intellectual property rights allowing the owner of that right to exclude legally parallel imports through trademark, copyright, and/or patent laws. For instance, the United States has national exhaustion on the three property rights fields, but authorized dealers mainly use trademark and copyright laws to block parallel trade (Gallini and Hollis 1999). The other extreme is an *international* exhaustion regime, which makes parallel imports legal. This is the case in Japan (with restrictions) and Australia and New Zealand (for most copyrighted and trademarks goods). A *regional* exhaustion regime, as in the European Union (EU), represents an intermediate case, since it makes parallel imports legal when they originate from EU members but illegal when they originate from non-EU members.

The main economic policy question about parallel imports is whether a property right owner should be entitled to international exclusive distribution territories. The fact that different regimes have been adopted indicates the existence of conflicting views about the motives and economic effects of parallel imports. In a nutshell, views range from banning parallel trade because intellectual property owners should have the right to control the international distribution of their products, to unregulated parallel trade because restrictions are nontariff barriers to trade that are inconsistent with World Trade Organization principles.

Price discrimination is generally considered the primary motive behind a manufacturer's support of a parallel import ban. Income differences, for instance, are typically more important across countries than within them, and they imply different demand elasticities across countries. This provides a strong incentive to manufacturers with market power to set country-specific prices. Because of arbitrage, this strategy would not be sustainable with parallel imports. Allowing international price discrimination (i.e., banning parallel imports) does not imply that international welfare decreases (Malueg and Schwartz 1994). A necessary condition for global

welfare to be higher under international price discrimination is that overall output be greater under discrimination than without it. This is a poor consolation for the high-price country but not for the low-price country, especially if the latter country is not served under a regime allowing parallel imports. This may well arise when income differences are large enough, a case that is relevant in the ongoing dispute about the availability of some critical pharmaceutical products in developing countries. In short, price discrimination does not provide a strong guide about policy prescription concerning parallel trade.

A second important motive behind manufacturers' attitude toward parallel trade is linked to vertical control over intermediaries selling these products to consumers. In particular, banning parallel trade and having exclusive intermediaries may help facilitating collusion among manufacturers or among dealers. It may also provide incentives to these exclusive intermediaries to undertake specific investments in activities such as advertising or after-sales services. If parallel imports were allowed, free riding would take place and, with it, very little incentive to invest in such activities. In other words, whenever exclusive territories are useful to motivate dealers, there are good reasons to ban parallel imports as well. Still, manufacturers who deal with intermediaries may have a strong incentive to let dealers engage in parallel trade. This is particularly the case in markets where production and shipments must take place before demand is really known and for products whose sale value drops at the end of the demand period (Raff and Schmitt 2007). If parallel trade was banned in such markets (for instance, confectionary), intermediaries could be stuck with worthless inventories when the demand in their market turns out to be low. Fearing this possibility, intermediaries would naturally place smaller orders to manufacturers than they would if they were free to resell unsold inventories in high-demand markets. Thus, when seasonality and uncertainty are important market demand features, manufacturers' profits may well be significantly higher when parallel trade is allowed than when it is not.

A third set of motives for banning parallel trade has to do with its dynamic impacts on markets. For instance, banning parallel trade may prevent confusion among consumers when the products have different specifications (whether due to different tastes across locations, safety standards, or packaging). More significantly, parallel imports may reduce manufacturers' incentives to engage in research and development or product innovation. This argument is often made in relation to the pharmaceutical industry (see Li and Maskus 2006).

These possibilities show that parallel trade can have nontrivial implications for the competitive environment of manufacturers and intermediaries. Clearly, a policy about parallel imports, whether at the national or at the international level, can rarely be reduced to a trade policy recommendation but must typically also include strong competition policy considerations.

Empirical evidence shows that the threat of parallel imports does decrease retail prices. In itself, these price effects are not enough to justify a free regime concerning parallel trade. Since a worldwide ban on parallel trade cannot be recommended either, a policy about parallel imports can be designed only on a case-by-case basis, depending on the characteristics of each country or region.

See also access to medicines; Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS); World Trade Organization

FURTHER READING

Gallini, Nancy, and Aidan Hollis. 1999. "A Contractual Approach to the Gray Market." *International Review of Law and Economics* 19 (1): 1 21. A very useful law and economics discussion of the U.S. intellectual property laws with respect to parallel trade.

Li, Changying, and Keith Maskus. 2006. "The Impact of Parallel Imports on Investments in Cost Reducing Re search and Development." *Journal of International Eco nomics* 68: 443 55. Develops a theoretical model about the links between innovations and parallel trade.

Malueg, David, and Marius Schwartz. 1994. "Parallel Im ports, Demand Dispersion, and International Price Discrimination." *Journal of International Economics* 37

(3 4): 167 95. The seminal theoretical paper on price discrimination and parallel trade.

Maskus, Keith. 2000. "Parallel Imports." *World Economy* 23 (9): 269 84. A nontechnical survey of the policy issues concerning parallel trade.

Raff, Horst, and Nicolas Schmitt. 2007. "Why Parallel Imports May Raise Producers' Profits." *Journal of In ternational Economics* 71 (2): 434 47. Develops a simple model explaining the circumstances under which al lowing parallel imports is beneficial to manufacturers.

NICOLAS SCHMITT

■ partial equilibrium models

Partial equilibrium models are used to analyze trade issues in a single market or, alternatively, in a few closely related markets. They are adaptations of standard supply and demand analysis to the specific features of trade policies. Partial equilibrium models are used in cases where linkages to other sectors of the economy are negligible enough to be ignored. They have the advantage of being relatively easy to construct and lend themselves to transparent application in policy analysis. They are extensively used in the analysis of anti-dumping and countervailing duties, as well as to assess the probable effects of many types of trade policy changes in narrowly defined sectors.

A decision not to use partial equilibrium modeling (or the more involved approach of applied general equilibrium modeling) usually results in the deployment of some sort of trend analysis, which focuses on whether the direction and timing of changes in one set of economic variables are coincident with another set of economic variables. Trend analysis is sometimes implemented in the form of an interpolation based on trend lines before and after a critical event, such as a change in trade policy. Trend analysis has two problems, however. First, the causative links between the two sets of variables are never made explicit, and the relative contribution of one factor to a given set of events is therefore difficult to assess. Second, while the number of necessary parameter estimates is minimal, this is simply a consequence of the implicit assumptions embedded in

the approach. Partial equilibrium analysis is a superior approach in almost all circumstances.

To construct a partial equilibrium trade policy model, the analyst first needs to determine whether the imported and domestic competing goods are perfect or imperfect substitutes and then needs to determine whether the country is "small" with reference to the rest of the world (in which case the import supply curve is horizontal or perfectly elastic) or "large" with reference to the rest of the world (in which case the import supply curve is upward sloping or less than perfectly elastic). These fundamental choices, which ideally should reflect the empirical reality of the situation being modeled, determine the appropriate modeling framework. If imports and domestic competing goods are best modeled as imperfect substitutes (following Armington 1969), the transmission of shocks from the market for the imported good to the market for domestic goods relies on the cross-price elasticity of demand or, alternatively, the elasticity of substitution. This measure affects the extent to which changes in the price of an imported good affect demand for the domestic competing good. Partial equilibrium models, whether perfect or imperfect substitutes, are typically implemented using spreadsheet software, examples of which are included in Francois and Hall (1997) and Roningen (1997).

A Perfect Substitutes Model We will first consider a perfect substitutes model with "second-best" effects in the "small" country case. This model proceeds as indicated in the supply and demand diagram of figure 1. The supply curve S represents the behavior of domestic firms, and the demand curve D represents the behavior of domestic households. Imports are available from the world market at a constant price P^w. It is the constant nature of this price that puts us in the "small" country case. The domestic government has imposed a specific tariff on imports of amount T_1. At the resulting domestic price of $P^w + T_1$, imports are of an amount equal to Z_1. The policy change we are to consider here is an increase in this tariff of up to level T_2.

The supply and demand curves in figure 1 are linear, that is, they can be expressed in the form of
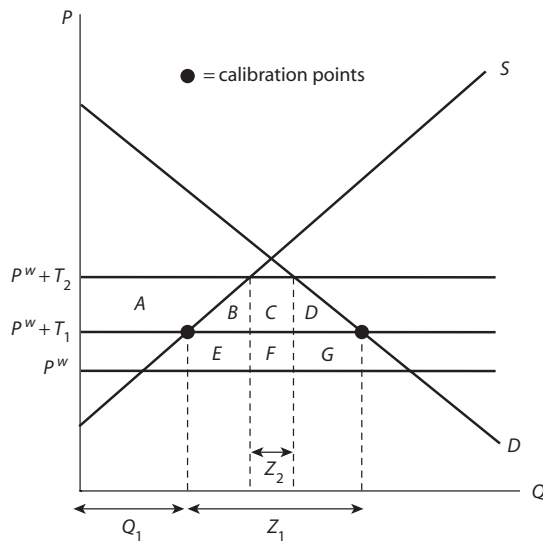
Figure 1
A small country, perfect substitutes model

$Q = a + bP$. However, it is also common to use a constant elasticity functional form, in which case the curves are expressed as $\ln Q = c + d \ln P$, where "ln" denotes the natural logarithm. In the former case, the price elasticities of demand and supply vary along the curves, while in the latter case, they are constant and equal to d.

The increase in the tariff level raises the domestic price to $P^w + T_2$ and reduces imports to Z_2. Consumer surplus (a measure of household welfare) falls by area $A + B + C + D$, and producer surplus (a measure of firm welfare) increases by area A. The government gains tariff revenue due to the increase of the tariff of area C but also loses tariff revenue due to the reduction in imports of areas E and G. Area F is collected as tariff revenue both before and after the increase in the tariff. The net result of all these changes is that welfare as conventionally measured declines by $B + D + E + G$. Note also that the increase in domestic output from Q_1 along the supply curve can be easily translated into an increase in employment using a fixed employment-output ratio and that this change is often of key political interest.

Calibrating this simple model involves the same three elements that are used in more complex mod-

els: functional forms, initial values, and elasticities. In the present context, the functional forms are those of the supply and demand curves (either linear or constant elasticity) and the elasticities are those related to these functions. The elasticities are estimated prior to applying the model. With regard to initial values, the calibration can be made easier by defining the initial domestic price ($P^w + T_1$) to be unity. Then the sum of the border value of the initial imports plus the initial tariff revenue gives the quantity Z_1, and the value of domestic output give the quantity Q_1. Q_1 determines the calibration point along the supply curve, and $Q_1 + Z_1$ determines the calibration point along the demand curve. This way of defining units is typical of applied trade policy models.

Imperfect Substitutes Model We next consider an imperfect substitutes model in the "large" country case, allowing for terms of trade effects. This model is presented in figure 2. The important difference between figures 1 and 2 is that, in the imperfect substitutes framework of the latter, there are now *two* closely related markets, one for the imported good Z and another for the domestic competing good D. The demand curves in these two markets are related through the cross-price elasticity of demand (or alternatively the elasticity of substitution) between the two goods. The initial equilibrium in the absence of a tariff results in the two prices P_{Z1} and P_{D1}. The imposition of a specific tariff T on imports of good Z causes the supply curve of this good to shift upward by the amount of the tariff, raising the domestic price of the imported good along the demand curve. The increase in the price of good Z affects the demand for good D, shifting the curve out as households substitute toward the domestic good. This increases the domestic price of good D and in turn causes a substitution toward good Z and a shift out of the demand curve for imports. These two substitution effects are simultaneous, and the resulting, new prices are P_{Z2} and P_{D2}.

We next consider the welfare effects of the tariff in this imperfect substitutes framework. In the market for the domestic good, there is an increase in producer surplus along the supply curve equal to trapezoid H (extending from the vertical price axis all the
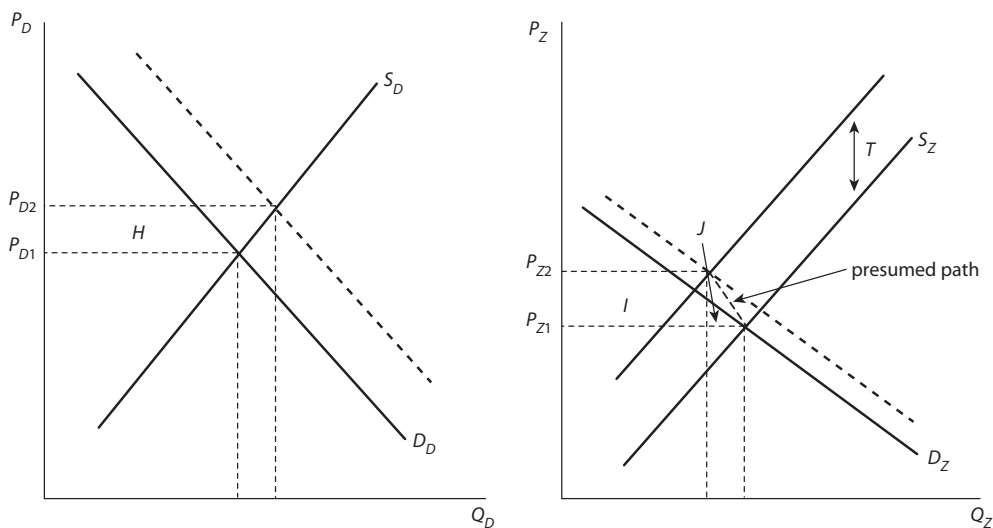
Figure 2
A large country, imperfect substitutes model

way to the supply curve). This entire area, however, comes as a cost to the consumers, with the producer gain and the consumer loss exactly offsetting each other. In the market of the imported good, there are no domestic producers to account for. However, the estimation of the consumer welfare effect is troubled by the fact that both the supply curve and the demand curve in the market for good Z have shifted. The standard approach to this, introduced by analysts such as Morkre and Tarr (1980), is to measure the change in consumer surplus along the presumed path between the initial and final equilibria points. The resulting consumer surplus loss is the trapezoid $I+J$. Rectangle I represents an increase in tariff revenue, so the net welfare effect in figure 2 is just triangle J.

One important aspect of figure 2 is that, discounting for the effect of the shift of the demand curve, the rise in the domestic price of the imported good is less than the tariff. This is because there is a movement in world quantity supplied down S_Z and a resulting decline in the border price of the imported good. This terms-of-trade effect of the tariff is missing in figure 1. The terms-of-trade effect has the property of reducing the height of the net welfare triangle J and is present unless the import supply

curve S_Z is horizontal or infinitely elastic, putting us back into a "small" country case.

As in the perfect substitutes model, the increase in domestic output along S_D can be easily translated into an increase in employment using a fixed employment-output ratio, and calibration to the initial equilibria points can be made easier by defining the initial domestic prices of the imported and domestic goods to be unity.

Partial equilibrium modeling of trade policy changes will continue to be an important analytical tool for situations where the linkages of the trade policy change to the broader economy under consideration are weak. Despite some complications in the imperfect substitutes case, the models are relatively easy to implement in spreadsheet form and are widely used by trade policy analysts.

See also applied general equilibrium models; gravity models; quotas; tariffs

FURTHER READING

Anderson, Kym. 1992. "The Standard Welfare Economics of Policies Affecting Trade and the Environment." In *The Greening of World Trade Issues*, edited by Kym Anderson and Richard Blackhurst. Ann Arbor, Michigan:

University of Michigan Press, 25 48. A noted application of partial equilibrium modeling to the issue of trade and the environment.

Armington, Paul. 1969. "A Theory of Demand for Products Distinguished by Place of Production." *IMF Staff Papers* 16 (3): 159 76. The key contribution on product differentiation by country of origin used in partial equilibrium trade policy modeling.

Baldwin, Robert E., and Tracy Murray. 1977. "MFN Tariff Reductions and Developing Country Trade under the GSP." *Economic Journal* 87 (345): 30 46. An early application of partial equilibrium trade policy modeling applied to the General System of Preferences using the imperfect substitutes model.

Francois, Joseph F., and H. Keith Hall. 1997. "Partial Equilibrium Modeling." In *Applied Methods for Trade Policy Analysis: A Handbook*, edited by Joseph F. Francois and Kenneth A. Reinert. Cambridge: Cambridge University Press, 122 55. A more advanced introduction to partial equilibrium trade policy modeling.

Morkre, Morris E., and David G. Tarr. 1980. *Effects of Restrictions on United States Imports: Five Case Studies and Theory*. Washington, DC: United States Federal Trade Commission. An early application of the imperfect substitutes model to U.S. trade policy.

Reinert, Kenneth A., and David W. Roland Holst. 1992. "Armington Elasticities for United States Manufacturing Sectors." *Journal of Policy Modeling* 4 (2): 631 39. The first estimation of elasticities of substitution between imported and domestic competing goods.

Roningen, Vernon O. 1997. "Multimarket, Multiregion Partial Equilibrium Modeling." In *Applied Methods for Trade Policy Analysis: A Handbook*, edited by Joseph F. Francois and Kenneth A. Reinert. Cambridge: Cambridge University Press, 231 99. An introduction to the extension of partial equilibrium trade policy modeling to multiple markets and multiple regions.

Rouslang, Donald J., and John W. Suomela. 1988. "Calculating the Welfare Costs of Import Restrictions in the Imperfect Substitutes Model." *Applied Economics* 20 (5): 691 700. An often cited study using the imperfect substitutes version of the partial equilibrium model.

KENNETH A. REINERT AND JOSEPH F. FRANCOIS

patents

See Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS); intellectual property rights

peso problem

The term *peso problem* generically refers to situations in which the possibility of a significant change in the distribution of future shocks to the economy influences the expectations of economic agents. The interaction of expectations and shifts in the distribution of economic shocks can lead to behavior in the prices of assets such as stocks and foreign exchange that is not in accord with the predictions of standard economic theory.

The precise origin of the term is uncertain, but is often attributed to comments made by Milton Friedman regarding the Mexican peso in the early 1970s. At that time, the Mexican peso traded at a fixed rate against the U.S. dollar while the interest rate on Mexican bank deposits exceeded that on comparable U.S. deposits. The existence of such an opportunity to earn a profit with little or no risk might appear to present a problem for rational-agent-based models of financial markets. Such an interest differential would fail to be arbitraged away, however, if investors thought that the peso would be devalued in the future. Indeed, in August 1976, the peso was allowed to float against the dollar and subsequently declined in value.

Peso Problems and Forecasts The peso problem is intimately related to economic forecasting. Good forecasts of future events have two important properties: they are unbiased, meaning that forecast errors are zero on average, and forecast errors are themselves unpredictable. Forecasting models are typically constructed and tested with historical data that are assumed to accurately represent the statistical distribution of the variables under consideration. When the statistical distribution is invariant over time, historical data can be used to assess the likelihood of future outcomes and generate forecasts with good properties. If the distribution of outcomes is shifting over time, however, then forecasts that are

conditioned on historical data may be biased or have predictable forecast errors because the historical data may not give an accurate representation of the changing statistical distribution.

Peso problems may occur when an economy faces this sort of instability. The use of historical data to predict the future becomes more difficult if the likelihood of future outcomes differs significantly from what was observed in the past. For example, wars and severe political turmoil are extremely hard to predict and may lead to significant changes in the economic environment. If economic agents believe there is a large enough possibility that such infrequent events might occur, there can be a significant impact on forecasts and forecast errors. To an outside observer, it may then appear that forecasters are making persistent errors, even though forecasters are using best practices in conditioning on such infrequent and hard-to-predict events.

Technically, peso problems can be interpreted as a failure of the methodology of rational expectations econometrics (Evans 1996). Rational expectations econometrics requires that the actual, or objective, distribution of economic variables be equal to the distribution expected by economic agents when they make decisions. The expected distribution may put weight on events that subsequently fail to be realized, and so may not be accurately represented in the ex post distribution. Under such circumstances, the predictions of rational expectations models may appear to be at variance with the data.

The Forward Premium Puzzle The forward premium puzzle is an asset-pricing anomaly that has lent itself to a peso problem interpretation. The foreign exchange market allows individuals to purchase forward contracts on currencies. A forward contract is an agreement to buy or sell a currency on a future date for a specified price called the forward rate. Consequently, the forward rate embodies investors' beliefs about the future value of the spot rate, which is the price at which a currency can be bought or sold for immediate delivery. The forward rate would thus seem to be a good predictor of the future spot rate on the date that the forward contract matures. A large amount of evidence indicates that the forward rate is not an unbiased predictor of the future spot rate, however. Under a peso problem interpretation, the bias, which is a tendency for the forward rate to stay above or below the spot rate for extended periods of time, could be due to investors' perceptions that there is a chance that the exchange rate will change substantially. If so, and until the exchange rate actually does shift, the forward rate will remain persistently above or below the spot exchange rate. Such a peso problem explanation of the forward premium puzzle has some validity in the data (Evans and Lewis 1995).

Stock Market Returns and Peso Problems Historically, the U.S. stock market has performed well relative to the stock markets of other industrialized nations. The historical, real return on U.S. equities is about 4.7 percent compared with a median return of about 1.5 percent for a sample of 39 countries (see Jorion and Goetzmann 1999). Peso problems offer one potential explanation for the large relative return. Investors must be rewarded for taking risks. Risk-averse stock market investors will demand a high return on equities in normal times to compensate them for the risk of extreme losses in bad times. But for the United States, extreme bad outcomes have largely bypassed the stock market. For example, unlike the stock markets of France, Russia, Germany, and Japan, the operation of the U.S. stock market was not interrupted for significant stretches of time by war. Nonetheless, stock market investors presumably attached some positive probability to a severe market disruption and factored that in to the returns they require to make their investments worthwhile. Since such a U.S. market disruption has not occurred, actual historical returns on equities have not reflected it. Thus the large realized U.S. equity return may be tied to investors' assessment of the possibility of extreme outcomes that never actually materialized for the U.S. economy.

Models Peso problems clearly present a difficulty for estimating and evaluating economic models. Good models require that events are realized often enough in historical data samples that researchers can accurately determine their probabilities of

occurrence. In the context of peso problems, this requires that infrequent events be observed often enough in the data for scholars to accurately assess the likelihood of repeated occurrence. One methodology that has been applied with some success to cases in which the distribution of outcomes shifts over time is the Markov regime-switching model (see Hamilton 1989). This model assumes that the economy can be in one of several distinct regimes and that regime switches are stochastic. That is, they change in a random fashion over time. If regime switches are observed to occur often enough in the historical data, researchers can estimate the probability of being in a particular regime and incorporate that information into the structure of economic models. This framework has had some success in modeling expectations of a possible future change in the economic environment that can in turn have implications for the behavior of asset prices.

See also capital mobility; exchange rate forecasting; foreign exchange intervention; forward premium puzzle; interest parity conditions; sovereign risk

FURTHER READING

Evans, Martin D. D. 1996. "Peso Problems: Their Theo retical and Empirical Implications." In *Statistical Methods in Finance*, edited by G. S. Maddala and C. R. Rao. Amsterdam: Elsevier, 613 46.

Evans, Martin D. D., and Karen Lewis. 1995. "Do Long Term Swings in the Dollar Affect Estimates of the Risk Premia?" *Review of Financial Studies* 8: 709 42. Both this and the preceding work by Evans study the ways in which peso problems can affect inference about foreign exchange risk premiums.

Hamilton, James D. 1989. "A New Approach to the Eco nomic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57: 357 84. Develops a methodology for incorporating Markov regime switch ing into models of economic time series. The technique has since been used in many papers that formally analyze peso problems.

Jorion, Philippe, and William Goetzmann. 1999. "Global Stock Markets in the Twentieth Century." *Journal of Finance* 54: 953 80. Collects data on equity returns for 39 markets going back to the 1920s and shows that U.S.

equities had the highest returns of all countries over 1921 to 1996.

Sill, Keith. 2000. "Understanding Asset Values: Stock Pri ces, Exchange Rates, and the 'Peso Problem.'" *Federal Reserve Bank of Philadelphia Business Review* (September/ October): 3 13. An introduction to how peso problems can affect the relationship between economic models and data.

KEITH SILL

■ petrodollars, recycling of

Recycling of petrodollars refers to what becomes of revenues earned by oil-exporting countries, particularly after major oil price shocks. Countries may spend some portion on foreign goods and services, but typically the increases in oil revenues following price shocks have been so large that oil exporters have been unable to absorb these petrodollars into their domestic economies either through consumption or through investment. Instead, petrodollars usually get recycled in the form of foreign asset savings held abroad. Such foreign assets may range from deposits held in foreign banks to bonds and private equity.

The most significant oil price shock followed the 1973 oil embargo imposed on the United States by the Organization of the Petroleum Exporting Countries (OPEC), but petrodollar recycling was one result of a steady increase in oil prices from 2002 to 2005. Oil prices reached an average $53 per barrel in 2005, more than twice their level just three years before. As a result, for a significant group of oil-exporting countries, oil export revenue rose from $262 billion in 2002 to an estimated $614 billion in 2005. These revenues in turn were invested in foreign assets, making oil exporters important counterparts to the United States in the ownership of foreign savings. In 2005 their current account surplus represented nearly 40 percent of the U.S. current account deficit.

What has the impact been and what is the impact likely to be of petrodollar recycling on global financial imbalances? If energy exporters spend their oil revenue, part of the petrodollar inflows would flow

out again to pay for imports of goods and services, boosting the economies of the industrialized oil-importing countries from which the petrodollars originated. This process would serve as a countervailing force to the oil price shock, rebalancing some portion of global imbalances. If, on the other hand, oil exporters save their revenue, this would exacerbate global imbalances by draining resources away from investment or consumption in the oil-importing world, in turn reducing growth and employment. To the extent that oil-exporting countries deposit their petrodollars in U.S. banks, however, this would mitigate the dampening effect. This is precisely what happened in the aftermath of 1973. But as we shall see, the real legacy of petrodollar recycling in the 1970s was not global financial imbalance but a third-world debt crisis.

Oil Shock of 1973 The oil shock of 1973 74 scaled up the size of postwar global economic imbalances by an order of magnitude and gave rise to widespread fears of a depression similar to the Great Depression of the 1930s. This episode gives important insights into issues of collective action problems, the resilience of international financial markets, and the uses and limitations of exchange rate adjustments.

The shock was generated by the large rise in oil prices following the oil embargo imposed against the United States in retaliation for its support of Israel in the Yom Kippur War of 1973. Because of the low price sensitivity of the demand for oil over the short and medium terms, this resulted in a significant increase in revenues for the oil-exporting countries. Combined with limitations on the speed with which the oil-exporting countries could increase their expenditures on imports (limitations of the capacities of their ports was a major initial bottleneck) and the fact that oil bills were paid primarily in dollars, this resulted in massive inflows of dollar reserves to oil-exporting countries, particularly in the Middle East.

The economies of the oil-importing countries were subject to three sets of pressures. Consumer prices rose while purchasing power drained away. For many oil-importing economies, the resulting stagflationary effects rising inflation combined with

reduced growth and employment were on the order of several percentage points of gross domestic product. On top of this, large current account deficits could cause severe balance of payments problems.

It soon became clear that in the aggregate there would not be a balance of payments problem corresponding to the huge current account imbalances. Had the world still been on a gold standard the situation would indeed have been dire since these imbalances would have drained gold out of the oil-importing countries, forcing monetary contraction and magnifying the initial recessionary effects of the oil price increases. Indeed, under the rules of the gold standard, the oil shock would have generated a new Great Depression.

Gold convertibility no longer existed, however, so the oil-exporting countries had to hold non-interest-bearing foreign currencies or, more reasonably, invest the surplus back into oil-importing countries through banks that accepted dollar deposits, primarily in Europe. Thus, in the aggregate, the oil surpluses created a financing glut instead of a financing shortage. In effect, the oil exporters recycled the funds by making very large deposits, which swelled the supply of loanable funds, forcing banks to offer extremely low-interest loans to any reasonable taker.

This in turn caused serious problems. Oil-importing countries with strong financial markets, such as the United Kingdom and the United States, began to run overall payment surpluses as the oil-related investment inflows exceeded their increased oil payments. As a result, there was an urgent need for secondary recycling of the petrodollars (as these surplus investment funds were called) more in line with the patterns of current account deficits.

Some of this alignment was accomplished by foreign aid from the oil exporters and some from a special recycling facility set up by the International Monetary Fund (IMF) in cooperation with oil-exporting nations. The oil exporters placed funds directly with the IMF that the IMF, in turn, then re-lent to oil-importing countries. The bulk of recycling took place through the international banking and financial sectors, however: a substantial amount of

the funds placed in the major financial centers (including the Eurodollar market) were re-lent mostly to newly industrializing countries (NICs) in Latin America and Asia.

Global Savings Glut Because of the sudden deposit glut, banks were eager to re-lend funds. The demand for loans within advanced industrial countries had slowed considerably due to recessionary conditions following the oil crisis. Banks saw the opportunity for big profits in loans to oil-exporting NICs, however, as oil prices had nearly quadrupled, making these countries appear to be good risks. Moreover, banks were attracted to the Latin American market, where authoritarian governments were standing ready to accept and guarantee repayment of private bank loans. Banks operated on the principle that sovereign loans carried very little risk of default compared with private sector loans. This was assumed true even for the oil-importing NICs, which also benefited from large loan packages.

Of course, Latin American countries exhibited considerable eagerness to borrow. Having reached the end of what scholars have referred to as "the easy phase" of import-substituting industrialization, Latin American NICs required new infusions of capital to maintain high growth rates. This process of borrowing to maintain high growth rates became known as debt-led growth. Moreover, the majority of Latin American NICs were governed by bureaucratic authoritarian regimes whose legitimacy hinged on their ability to continue delivering rapid industrial growth. To these regimes, the ability to borrow from private banks and thus bypass IMF conditionality was simply a bonus. But perhaps Latin American NICs' greatest incentive to borrow grew out of financial market conditions. Inflation had skyrocketed after the 1973 oil price hikes while nominal interest rates had leveled off and even fallen in some cases due to the credit glut. Together these phenomena resulted in extremely low and even negative real interest rates. Quite literally, for a time, banks were paying Latin American governments to take their money.

Another concern with respect to global financial imbalances after the oil shock was that instead of borrowing to finance their current account deficits,

countries would want to adjust their payments imbalances. But in the aggregate this was not possible in the short and medium term. Devaluations by all oil-importing countries would in effect raise the real price of oil, and because of the short-term inelasticity of demand for oil such actions would increase rather than reduce the aggregate current account imbalances.

For individual countries, however, devaluation would work by increasing their exports to, and reducing their imports from, other oil-importing countries. This process could work only if it were not practiced by too many other countries. This collective action problem was quite similar to the one countries faced during the Great Depression. Any one country could stimulate its domestic economy by increasing exports and reducing imports, but if all countries followed this strategy, the result would worsen the economic situation for everyone. In the 1930s, the combination of understanding of this problem and willingness to cooperate internationally was insufficient to avoid economic catastrophe.

Avoiding Beggar-Thy-Neighbor Policies Another Great Depression, therefore, was a realistic concern in the 1970s. Fortunately both the economic understanding of government officials and willingness to pay some attention to the international repercussions of national actions had developed sufficiently so that with valiant efforts from international organizations such as the IMF and the Organisation for Economic Co-operation and Development a repeat of the widespread beggar-thy-neighbor problems of the 1930s was avoided. Of course, cooperation was far from complete and many countries were criticized for not taking what was considered by others to be their fair share of the collective oil deficits (numerous calculations of "fair shares" were undertaken by various economists and organizations), but overall the initial handling of the oil shock was much better than could have been realistically hoped for.

Indeed, the international pressure to avoid overadjustments in the form of reduced spending and economic contraction following the oil shocks was so successful that over the medium term many developing countries did not adjust enough, continuing to

rely for too long on heavy amounts of borrowing. This, in turn, was one of the major contributing factors to the Latin American debt crisis of the 1980s. Nevertheless, the overall success in avoiding a repeat of the economic catastrophes of the 1930s was a major tribute to the growth of international cooperation and set the basis for dealing with several subsequent major run-ups in oil prices.

See also balance of payments; beggar-thy-neighbor policies; convertibility; Eurocurrencies; expenditure changing and expenditure switching; global imbalances; gold standard, international; International Monetary Fund (IMF); International Monetary Fund conditionality; international reserves; Latin American debt crisis; petroleum; twin deficits

FURTHER READING

Nsouli, Saleh. 2006. "Petrodollar Recycling and Global Imbalances." Presented at the CESifo's International Spring Conference. Berlin: Center for Economic Studies, Information and Forschung Institute (March 23 24). Downloadable from http://www.imf.org/external/np/speeches/2006/032306a.html. A useful analysis of petrodollar recycling in the new millennium.

Willett, Thomas D. 1995. "The Oil Transfer Problem and International Economic Stability." Princeton Essays in International Finance No. 113 (December). Princeton, NJ: Princeton University. Provides a more detailed treatment of the 1973 oil shock and its international financial repercussions.

THOMAS D. WILLETT AND NANCY NEIMAN AUERBACH

■ petroleum

The world petroleum sector includes the exploration, extraction, and transportation of crude oil, and the refining of crude oil into finished products. Participants in the world petroleum sector include fully integrated oil companies, national oil companies, independent oil and gas producers, refiners and marketers, pipeline operators, and others. The fully integrated companies explore for and produce oil and gas around the world, own pipelines and tankers to transport oil and gas, process the crude oil into refined products, and sell finished products through a global network of wholesale and retail outlets. Typical fully integrated companies are often international oil companies (IOC) such as Exxon Mobil Corporation, BP, Chevron Texaco, ConocoPhillips, and Royal Dutch/Shell. In contrast to IOCs, national oil companies are owned by national governments and are typically found in major oil-producing nations. Independent oil and gas producers only explore and/or produce crude oil and natural gas. Independent refiners purchase crude oil and process it into finished products. These companies may also own wholesale and retail marketing outlets or sell their products to marketing companies. Independent marketers purchase refined products, usually gasoline, and sell them at retail outlets. Pipeline companies transport crude oil, refined products, natural gas and natural gas liquids using networks of pipes and pumping or compressor stations.

Oil and natural gas provide nearly 60 percent of the world's primary energy and will remain indispensible in meeting the projected growth in energy demand during the early decades of the 21st century. The rapidly growing world economy will require large increases in oil, which will come from a variety of sources, including existing production capacities, development of existing reserves, new discoveries, and development of nonconventional liquids. Yet there is uncertainty about the industry's ability to overcome the multiple increasing risks of meeting the growing demand for oil. Such risks include gaining access to promising resources, making investment in infrastructure, and determining how much oil is recoverable.

Oil Demand Between 1960 and 1972, world consumption of oil increased by one and one-half times, about 7 percent on average per year. In turn, the world's industries, including transportation and commerce, and individual households became increasingly dependent on oil. In the years immediately following the Arab oil embargo in 1973 and the subsequent shock in 1979, world demand for oil greatly diminished. From 1980 to 2000, the annual growth in world oil demand averaged 0.9 percent.

The average growth in world oil demand for the future is expected to be much greater than in the past. Between 2000 and 2030, oil demand is projected to increase at an annual rate of from 1 percent to 1.9 percent. World demand for oil is expected to grow from about 76 million barrels per day in 2000 to between 98 and 138 million barrels per day in 2030. On a global basis, the transportation sector accounts for 68 percent of the total projected increase in oil consumption between 2004 and 2030, followed by the industrial sector, which accounts for another 27 percent of the increased consumption (Energy Information Administration, *International Energy Outlook, 2007*).

The largest increases in oil consumption from 2004 to 2030 are projected to be 7 million barrels per day in North America and 15 million barrels per day in Asian countries that do not belong to the Organisation for Economic Co-operation and Development (OECD). It is projected that non-OECD countries' oil consumption will continue to grow, driven by strong economic and industrial growth and rapidly expanding transportation use. It is further estimated that the fastest growth in oil consumption will occur in non-OECD Asia, averaging 2.7 percent per year from 2004 to 2030. Non-OECD Asia accounts for 43 percent of the overall increase in world liquids consumption (Energy Information Administration, *International Energy Outlook 2007*).

Prices Historically, world oil prices were low and stable in the 1950s and 1960s, driven by new discoveries in the Middle East, the North Sea, Alaska, and Nigeria. Cheap oil thus contributed to strong worldwide economic growth, which in turn stimulated greater consumption. With the first oil shock in 1973, prices increased fourfold, and in 1979, prices increased by another threefold. Prices thereafter declined haltingly and gradually until 1986 and then collapsed until late 1992. From 1992 to the beginning of 2007, oil prices increased fivefold.

Future projections of world oil prices, in constant 2006 dollars, have been projected by the U.S. Energy Information Administration (EIA) to decline from $68 per barrel in 2006 to $49 per barrel in 2014, then rise to $59 per barrel in 2030 ($95 per barrel on a nominal, noninflation-adjusted basis). The EIA low- and high-price cases suggest great uncertainties in the reference case projections, however. In the low-price case, world oil prices are projected to be $36 per barrel in 2030 ($58 per barrel on a nominal basis), while in the high-price case, oil prices are estimated to be $100 per barrel in 2030 ($157 per barrel on a nominal basis).

Production Oil production is projected to increase by 35 million barrels per day in 2030 over 2004 levels. To meet the demand for oil, production is expected to increase in both the Organization of the Petroleum Exporting Countries (OPEC) and non-OPEC producers, with most (65 percent) of the total increase coming from OPEC countries.

It is generally acknowledged that OPEC members with large reserves and relatively low costs for expanding production capacity can accommodate sizable increases in the world's petroleum consumption. Geopolitical issues in a number of the OPEC countries, including Iraq, Iran, Venezuela, and Nigeria, lead to great uncertainty in making projections of future production levels. Projections of Iraq's future oil production are well below its prewar production.

Reserves and Resources As of January 1, 2007, "proved" world oil reserves were estimated at 1,208 billion barrels. Proved reserves are estimates of the amount of oil recoverable from known reservoirs under current economic and operating conditions. Proved reserves reflect only a fraction of the oil that a reservoir may hold. "Recoverable" oil includes proved reserves and undiscovered oil that is economically extractable.

Reserves are concentrated in the Middle East and North Africa (MENA), together accounting for 62 percent of the world total. Saudi Arabia, with the largest reserves of any country, holds about a fifth of the world's total reserves. Of the twenty countries with the largest reserves, seven are in the MENA region. Iranian oil reserves have increased by 46.6 billion barrels, or 52 percent, since 2000. Kazakhstan has had the third-largest increase, 24.6 billion barrels, since 2000. Among the top 20 reserve holders in

2007, 11 are OPEC member countries that, together, account for 65 percent of the world's total reserves.

The United States Geological Survey (USGS) mean estimate of ultimately recoverable global conventional oil plus natural gas liquids (NGL) was 3.345 trillion barrels at the beginning of 1996. The estimates range from 2.5 to 4.4 trillion barrels, expressed in statistical terms, where 2.5 trillion barrels represents a 95 percent probability that the quantity size will exceed the estimate, and 4.4 trillion barrels indicates a 5 percent probability that the quantity will exceed the estimate. By comparison, IOCs project an average of 3.5 trillion barrels. The IOC most-likely estimates for ultimately recoverable global conventional oil range from 2.8 to 4.0 trillion barrels.

The uncertainty of ultimate recoverable oil and NGL has led to a continuing debate about the prospect of a future oil production peak and eventual decline. The timing of a peak and subsequent decline is subject to interpretation because of a number of variables, such as lack of transparent reporting of producing fields, conflicting estimates of recoverable oil, the increased costs of producing oil from conventional and nonconventional sources, and the timing and scale of development of alternative fuels.

Oil Companies State-owned, or national, oil companies control most of the global proved oil reserves. As of 2005, the global proved oil reserves were 1.148 trillion barrels, with national oil companies (NOCs) in control of 77 percent of the total (886 billion barrels) and allowing no private equity ownership. According to an 2005 annual survey published by *Petroleum Intelligence Weekly*, 12 of the top 20 oil companies are traditional (100 percent state-owned) national oil companies (NOCs) or hybrid, partly state-owned NOCs. *Petroleum Intelligence Weekly*'s ranking shows that Saudi Aramco, Iran's NIOC, Iraq's INOC, Kuwait's KPC, Venezuela's PDVSA, United Arab Emirates' Adnoc, Libya's NOC, and Nigeria's NNPC hold the largest reserves in the world.

National oil companies' objectives often differ greatly from those of the privately owned international oil companies. The NOC objectives go beyond maximizing returns on investments, and often include redistribution of wealth in society, foreign policy objectives, energy security, wealth creation, and economic development. The NOCs' noncommercial objectives tend to interfere with decisions regarding investments in oil production expansion. However, not all NOCs have the same interests. In more mature NOCs, such as Statoil (Norway) and Petronas (Malaysia), the focus is on returns on investment, and the companies exhibit more familiar corporate behavior. At the opposite extreme, the business practices of PDVSA (Venezuela) and NNPC (Nigeria) are heavily influenced by domestic policies. These firms are focused squarely on increasing the revenues to the government. Saudi Arabia has used its oil reserves in its foreign policy strategy, given its role as the world's oil swing producer. In its role as having the largest spare crude oil production capacity in the world, Saudi Arabia has the ability to replace the exports of any small- or medium-sized oil-producing country with days or weeks.

Investment The International Energy Agency estimates that over the period of 2005–2030 $4.3 trillion in new investments will be needed in the global oil sector to meet rising world demand for oil. The largest investment projections are in the Middle East and developing Asia, where most of the upstream (exploration and production) and downstream (refineries, pipelines, tankers) investments will be needed. The upstream sector accounts for the bulk of this investment requirement, with required investments averaging $125 billion per year. Almost three-quarters of the upstream investments will be required to maintain existing production capacity in the face of declining capacity of existing oil wells as reserves are depleted. Downstream investment requirements are projected to be about $700 billion from 2005 to 2030. This projection includes the addition of new refineries to meet the rising demand and additional investments on existing refineries.

There is no guarantee that the projected levels of investment will actually occur, as several barriers exist to upstream and downstream investments. First, most privately owned IOCs have large cash reserves and are able to borrow at reasonable rates from

petroleum

capital markets when necessary for new projects. Because of restrictions on access to oil reserves in many resource-rich countries, however, many IOCs may not be able to invest in upstream development. A large portion of oil reserves are found in countries where there are restrictions on foreign investment. Kuwait, Mexico, and Saudi Arabia, for example, remain totally closed to upstream oil investment by foreign companies. Other major oil-producing countries, such as Venezuela, Russia, Iran, Algeria, and Qatar, have effectively restricted foreign investment as well. Second, even in those cases where IOCs are willing to invest, other barriers such as licensing and fiscal terms, or the investment climate, may discourage them from doing so. Stability of the political regime, war or civil conflict, or other geopolitical tensions act to dissuade inward investment in the development of countries' resources.

See also Organization of the Petroleum Exporting Countries (OPEC); petrodollars, recycling of

FURTHER READING

Baker Institute for Public Policy. 2007. "The Changing Role of National Oil Companies in International Energy Markets." March. Houston, TX: Baker Institute. Provides a review of the strategies, objectives, and performance of national oil companies.

BP Statistical Review of World Energy. 2007. http://www.bp.com/liveassets/bp internet/globalbp/globalbp uk english/reports and publications/statistical energy review 2007/STAGING/local assets/downloads/pdf/statistical review of world energy full report 2007.pdf. June. Authoritative source of data related to demand and production of oil and natural gas by region and country.

Energy Information Administration. 2007. *International Energy Outlook 2007*. Washington, DC: EIA. Provides an assessment of the outlook for international energy markets through 2030.

International Energy Agency. 2006. *World Energy Outlook 2006*. Paris: OECD/IEA. Provides an assessment of the outlook for international energy markets and additional insights into the most critical energy issues.

National Petroleum Council. 2007. *Hard Truths Facing the Hard Truths about Energy*. Washington, DC: NPC.

Evaluates the ability of global oil and natural gas supply to keep pace with growing world demand.

Simmons, Matthew R. 2005. *Twilight in the Desert The Coming Saudi Oil Shock and the World Economy*. New York: Wiley. Analyzes Saudi Arabia's oil exploration and production industry and whether it will be able to continue in a swing producer role in the coming decades.

Yergin, Daniel. 1991. *The Prize The Epic Quest for Oil, Money, and Power*. New York: Simon and Schuster. Tells the history of oil and how it has shaped the politics and the global economy.

MICHAEL R. CURTIS

■ pharmaceuticals

The pharmaceutical industry provides an important part of fundamental health care to the citizens of any given nation. Besides its social value, the industry is also important in economic terms its innovative capabilities contribute significantly to the knowledge and skills-based economy of advanced industrialized nations. Since the 1970s, the pharmaceutical industry has experienced considerable structural change. Government regulation continues to play a significant role in the industry, as does the drug discovery process and the cost of bringing a new drug to market.

Government Regulation The majority of new drugs are developed within the United States, the European Union (EU), and Japan, where the agencies responsible for their approval are the Food and Drug Administration; the European Medicines Agency (EMEA); and the Ministry of Health, Labor, and Welfare, respectively. Within the EU, marketing authorizations can be granted by either the EMEA or individual member states. If the product is derived from biotechnology, a firm must use the centralized system of the EMEA. Otherwise, if the firm chooses the national route, the procedure of mutual recognition, in which a product that has been judged safe for sale in one member state will be sold in all other member states, is adopted.

Although government regulation has increased to ensure the efficacy and safety of new drugs, important steps have also been taken toward increased international harmonization. The EU, for example, developed a single market for pharmaceuticals, culminating in the formation of the EMEA. In addition, the International Conference on Harmonization (ICH) was set up in 1990 to bring together regulatory authorities and drug developers from the EU, the United States, and Japan to attempt to eliminate duplicative requirements for drug development and approval. Since its inception, ICH has achieved the harmonization of technical guidelines and, more recently, of the format and content of registration applications.

Variations in pharmaceutical pricing, however, remain extreme due to the many differences in national health care systems. In Japan and the EU member states, the government is the main purchaser, and indeed, parallel trade (estimated at 4.2 billion euros in 2005) is allowed under the single-market rules. Thus third parties exploit price differences by buying pharmaceuticals in markets where low prices are enforced and selling them to governments and other purchasers in markets where higher prices have been agreed. In the United States, although the government is not the main purchaser, the rapid expansion of health care maintenance organizations has led to a concentration of purchasing power.

The only major national markets in which firms are free to set the price of new drugs are the United States, the United Kingdom, and Germany (since 2005), but soaring health care costs in advanced industrialized nations have led to tougher price competition. Governments and other purchasers have reduced reimbursement rates, increased the use of "limited lists" (which define drugs that may be prescribed and/or reimbursed for different diseases), and placed pressure on prescribers to substitute generics for branded drugs wherever possible. For example, the generic market grew by 13 percent in the eight largest national markets in 2005, compared to around 5 percent for "branded" pharmaceuticals, and generic prescription volume was greater than branded volume for the first time ever in the United States.

Drug Discovery Process At the end of World War II, the pharmaceutical industry experienced a radical shift from very little new drug development toward a research-oriented environment in which firms essentially screened thousands of chemical compounds for efficacy against a given disease (so-called random screening). Serendipity mattered: the mechanisms of how drugs worked were not well understood, and firms correspondingly maintained huge libraries of chemical compounds. As basic biomedical knowledge increased in the 1970s, a second technological shift occurred: "rational drug design" was adopted, which involved more precise models of how particular diseases function and the design of molecules that target particular cells or cause particular biological interactions. Since the 1980s, biologics (derived from biotechnologies) constitute the third shift and provide a crucial arena for the development of new drugs in the 21st century. Approximately 20 percent of new drugs launched on the world market are biotechnology derived, and biologic sales were $52.7 billion in 2005. The United States was the single most important national market in 2005 in terms of research and development (R&D) expenditure (78 percent) and turnover (76 percent), with approximately 140,000 employees as compared with Europe's 33,000.

Over time, the average cost of developing a new drug has increased substantially. The cost was estimated at $897 million in 2003, compared with $500 million in the 1990s and $231 million in 1987. This compares with the first ever estimate of the average cost of developing a new biologic, which was estimated to be $1.2 billion in 2006, reflecting both the cost of time and the drugs that fail (Tufts Center for the Study of Drug Development). This increased cost is primarily due to the increasing focus on chronic rather than acute illnesses as populations age (i.e., targeted diseases such as Alzheimer's are more "complex," and the cost of R&D varies across therapeutic groups), tougher governmental regulations that have lengthened development times and increased the number of clinical trials, and finally, the need to convince prescribers that the new drug has significant benefits as compared with rival pharmaceuticals on the market.

Table 1
Top 20 global market shares and R&D expenditure

	2005				1998		1992	
	Sales ($bn)	Market share (%)	R&D ($bn)	Top selling drugs ($bn)	Market share (%)	R&D ($bn)	Market share (%)	R&D ($bn)
Pfizer (US)	44.3	7.4	7.4	Lipitor: 12.2 Norvasc: 4.7 Zoloft: 3.3	3.9	2.3	2.0	0.8
GlaxoSmithKline (UK)	34.0	5.6	5.7	Advair: 5.5 Avandia: 2.1 Lamictal: 1.5	4.2	1.9	3.8	1.0
Sanofi Aventis (Fr)	33.9	5.6	5.0	Lovenox: 2.7 Plavix: 2.5 Taxotere: 1.9	2.5	1.4	2.6	0.8
Novartis (Ch)	25.0	4.2	4.4	Diovan: 3.7 Gleevec: 2.2 Zometa: 1.2	4.2	1.8	2.2	0.8
Astra Zeneca (UK)	24.0	4.0	3.4	Nexium: 4.6 Seroquel: 2.8 Seloken: 1.7	2.8	1.3	1.4	0.4
Johnson & Johnson (US)	22.3	3.7	4.4	Risperdal: 3.6 Eprex: 3.3 Remicade: 2.5	3.6	1.8	1.9	0.9
Merck (US)	22.0	3.7	3.8	Zocor: 4.4 Fosamax: 3.2 Cozaar: 3.0	4.2	1.8	3.6	1.1
Roche (Ch)[a]	21.9	3.6	4.0	Rituxan: 3.4 Epogin: 1.8 Herceptin: 1.7	3.0	1.9	2.1	1.1
Wyeth (US)	15.3	2.5	2.6	Effexor: 2.5 Protonix: 1.7 Prevnar: 1.5	3.1	1.5	2.0	0.6
Bristol Myers Squibb (US)	15.3	2.5	2.5	Plavix: 3.8 Pravachol: 2.3 Avapro: 1.0	3.9	1.6	2.8	1.1
Eli Lilly (US)	14.7	2.4	3.0	Zyprexa: 4.2 Gemzar: 1.3 Humalog: 1.2	2.9	1.7	2.0	0.9
Abbott (US)	13.7	2.3	1.8	Humira: 1.4 Depakote: 1.0	2.5	1.2	1.8	0.8
Amgen (US)	12.0	2.0	2.3	Neupogen: 3.5 Aranesp: 3.3 Enbrel: 2.6	X	X	X	X
Boehringer Ingelheim (Ger)	11.4	1.9	1.7	Spiriva: 1.2 Mobic: 1.1	1.4	0.9	1.3	0.4
Takeda (Jap)	9.0	1.7	1.5	Prevacid: 3.4 Actos: 2.2 Blopress: 1.7	1.6	0.6	1.5	0.5

Table 1

(*continued*)

	2005				1998		1992	
	Sales ($bn)	Market share (%)	R&D ($bn)	Top selling drugs ($bn)	Market share (%)	R&D ($bn)	Market share (%)	R&D ($bn)
Astellas (Jap)[b]	7.5	1.2	1.2	Prograf: 1.2 Harnal: 1.2	X	X	X	X
Schering Plough (US)	7.6	1.3	1.9	Remicade: 0.9	2.5	1.0	1.5	0.5
Novo Nordisk (Den)	5.6	0.9	0.8	Antidiabetics: 4.0	X	X	X	X
Bayer (Ger)	5.0	0.8	1.2	Kogenate: 0.8	2.1	1.1	1.8	0.9
Schering (Ger)	4.8	0.8	1.2	Betaferon: 1.1	X	X	X	X
Total Sales ($bn)		602.0	66.3		302.0	40.0	229.9	26.0
TOP 5 SHARE		26.8	40.6		20.4	24.3	15.0	20.8
TOP 10 SHARE		41.6	65.9		35.9	44.7	25.4	38.0

Source: Firms' Annual Reports, *Pharmaceutical Executive* (May 2006).

Note: Where merger/acquisition has occurred, the market shares in 1998 and 1992 represent the larger of the two previously separate entries. Thus, for GlaxoSmithKline, the 1998 and 1992 market share data represent Glaxo's only.

[a] The data for Roche include Genentech's sales (as they own a majority holding). Genentech's sales in 2005 = $5.5 billion, with $1.3 billion on R&D.

[b] Astellas was incorporated on April 1, 2005, through the merger of Yamanouchi (Jap) and Fujisawa (Jap).

X: outside the Top Twenty

On average, it takes 12 years from a scientific discovery to bring the drug to market. The shorter effective patent life (reflecting the longer development period, which leaves a shorter period for sales of the pharmaceutical before it goes off patent), the increased ability by rival firms to use rational drug design to produce close substitutes, and the increase in generic competition due to changes in regulations have all eroded the length of market protection once enjoyed by the pioneer developer of a pharmaceutical. As the competitive environment has moved from the national to the global level, firms currently attempt to launch a new drug in as many major national markets as possible, as well as design various (incrementally innovative) formulations of the drug, license promising new compounds, and advertise directly to consumers where government regulations allow.

Industry Structure The pharmaceutical industry is best characterized as a global oligopoly in which large multinational firms dominate. Moreover within these firms, "blockbusters" (defined as drugs that

generate more than $1 billion in worldwide sales annually) dominate the product range, as shown in table 1. A good descriptive measure of success is the proportion of leading global medicines that national firms (in terms of the location of company headquarters) account for. In 2004, Germany owned 0 of the top 75 (as measured by worldwide sales), France owned 5, Japan owned 5, Switzerland owned 8, the United Kingdom owned 17, and the United States owned 40.

Total global pharmaceutical sales in 2005 were $602 billion, with the United States accounting for 44 percent, the EU for 30 percent, and Japan for 11 percent. Table 2 highlights the substantial increase in merger and acquisition activity in the 1990s, where a significant proportion was cross-border activity. Table 1 also gives the market shares of the 20 leading firms in the period 1992 2005, showing that the ranking of the leading companies changed significantly even in this rather short time period. Note also the relatively high increase in industry concentration. In 1992, the top five firms accounted for 12.9 percent of the global market. By 1998 this had increased to

Table 2

Merger and acquisition activity in the pharmaceutical industry

1989	SmithKline Beckman (US) and Beecham (UK)
	Bristol Myers (US) and Squibb (US)
	Dow (Merrell) (US) and Marion (US)
	American Home Products (AHP) (US) and AH Robins (US)
1990	Rhône Poulenc (Fr) and Rorer (US)
	Roche (US) and Genentech (US:60%)
1993	Synergen (US) and Amgen (US)
	Hoechst (Ger) and Copley (US)
1994	Ciba Geigy (Ch) and Chiron (US: 50%)
	AHP (US) and American Cyanamid (US)
	Roche (Ch) and Syntex (US)
	SmithKline Beecham (UK) and Sterling Health (US) (resold the OTC part to Bayer [Ger])
1995	Glaxo (UK) and Wellcome (UK); Glaxo (UK) and Affymax (NL)
	Hoechst (Ger) and Marion Merrell Dow (US)
	Pharmacia (Swed) and Upjohn (US)
	Rhône Poulenc Rorer (Fr) and Fisons (UK)
	BASF (Ger) and Boots (UK)
1996	Ciba Geigy (Ch) and Sandoz (Ch)
1997	Roche (Ch) and Boehringer Mannheim (Ger)
1998	Astra (Sweden) and Zeneca (UK)
	Rhône Poulenc Rorer (Fr) and Hoechst Marion Roussel (Ger)
	Sanofi (Fr) and Synthélabo (Fr)
1999	Monsanto (US) and Pharmacia & Upjohn (US/Sweden)
2000	G. D. Searle (US) and Pharmacia & Upjohn (US/Sweden)
	Pfizer (US) and Warner Lambert (US)
	Glaxo Wellcome (UK) and SmithKline Beecham (UK)
	Abbott Laboratories (US) and BASF Knoll (Ger)
2001	Bristol Myers Squibb (US) and DuPont Pharmaceuticals (US)
2003	Pfizer (US) and Pharmacia Upjohn (US/Sweden)
2004	Aventis (Fr) and Sanofi (Fr)
2006	Bayer (Ger) and Schering (Ger)

20.4 percent and to 26.8 percent by the end of 2005. The leading firms also rapidly increased their proportion of industry R&D: in 2005, for example, although the 10 top firms accounted for only 41.6 percent of the market by sales, they accounted for 65.9 percent of total R&D expenditure. Moreover it must be noted that an increasing proportion of firm-level R&D was spent in the United States, at the expense of Europe and Japan. In 1990, the respective R&D expenditures were $6.8 billion, $9.9 billion, and $3.6 billion, compared with $29.6 billion, $26.3 billion and $8.4 billion, respectively, by 2004.

Outlook Health care costs have increased steadily in advanced industrialized nations, due mainly to aging populations. In turn, this has led to the increased substitution of drugs for more expensive alternatives such as surgery and hospitalization. Moreover rapidly rising health care costs have led to more intense pressure on how pharmaceuticals are priced. National governments have tried various containment measures, such as allowing increased

competition from generics. In addition, R&D costs have risen due to the increasing costs of clinical trials and other regulatory hurdles, as well as the focus on chronic rather than acute diseases and the movement toward new drug discovery through biotechnology. In response, the industry experienced unprecedented mergers and acquisitions during the 1990s.

Overall the pharmaceutical industry plays a vital role in the modern world economy. Health care is one of the major issues when thinking about what welfare means in the 21st century. The industry is an enormous generator of innovation, investment, and highskilled employment. Given the high cost of innovation, notwithstanding the significant public funding in basic medical research, however, there remains a relative lack of research into rare (genetic) diseases that afflict the few and diseases that primarily affect developing countries. Other major issues for the industry include the protection of intellectual property and the improvement of access to medicines in developing countries.

See also access to medicines; foreign direct investment under oligopoly; health and globalization; mergers and acquisitions; multinational enterprises

FURTHER READING

Association of the British Pharmaceutical Industry. 2006. *Pharmaceutical Industry Competitiveness Taskforce: Competitive and Performance Indicators 2005.* Available at http://www.abpi.org.uk. Trade association that provides industry statistics and important information and articles, primarily on the UK pharmaceutical industry.

Busfield, J. 2003. "Globalization and the Pharmaceutical Industry Revisited." *International Journal of Health Services* 33 (3): 581 605. A survey of the trends in the pharmaceutical industry at the beginning of the 21st century.

European Federation of Pharmaceutical Industries and Associations. 2006. *The Pharmaceutical Industry in Figures.* Available at http://www.efpia.org. Brussels. Represents the research based pharmaceutical industry in Europe; provides information/articles and statistics.

Gambardella, A., L. Orsenigo, and F. Pammolli. 2000. *Global Competitiveness in Pharmaceuticals: A European Perspective.* Brussels: European Commission Enterprise

and Industry Directorate General. An empirical investigation into the determinants of firm performance in the pharmaceutical industry.

German Association of Research Based Pharmaceutical Companies. Industry statistics and other information available at http://www.vfa.de/en/articles/index en.html. Trade association that provides industry statistics and important information and articles, primarily on the German pharmaceutical industry.

Helms, R. B., ed. 1996. *Competitive Strategies in the Pharmaceutical Industry.* Washington, DC: The American Enterprise Institute Press. Notable authors consider pharmaceutical pricing, forces affecting global competition, and the risks of and returns to pharmaceutical research and development.

Matraves, C. 1999. "Market Structure, R&D, and Advertising in the Pharmaceutical Industry." *Journal of Industrial Economics* 47 (2): 169 94. An empirical investigation of the relationship between sunk costs and market structure.

Pharmaceutical Executive. Industry statistics and other information available at http://www.pharmexec.com/pharmexec/. Industry publication that covers key trends and issues, as well as providing industry data.

PhRMA. Industry statistics and other information available at http://www.phrma.org/. Trade association that provides industry statistics and important information and articles, primarily on the U.S. pharmaceutical industry.

U.S. Congressional Budget Office. 2006. *Research and Development in the Pharmaceutical Industry.* Washington, DC: CBO. An overview of pharmaceutical R&D expenditure and new drug output.

CATHERINE MATRAVES

■ Plaza Accord

The Plaza Accord was an agreement reached on September 22, 1985, at the Plaza Hotel in New York, among the G5 countries (the United States, Japan, Germany, France, and the United Kingdom) to push down the value of the dollar in order to reduce large U.S. trade deficits. On the day that the Plaza Accord was announced, the trade-weighted or effective value

of the dollar fell 4 percent. The dollar fell 17 percent against the Japanese yen and 15 percent against the German mark between September 1985 and the end of the year.

Background In the 1970s, the United States had frequently sought to push down the value of the dollar to reduce its trade deficit. In 1976 and 1977 expansionary monetary and fiscal policies, together with attempts by Treasury secretary Michael Blumenthal to "talk down" the dollar, caused the exchange rate to depreciate (Frankel 1994). The fall of the dollar, aggravated by inflationary pressures, subsequently led to consternation among policymakers.

In October 1979 Fed chairman Paul Volcker initiated a change in monetary policy operating procedures designed to fight inflation and strengthen the dollar. The new strategy allowed interest rates to reach very high levels if necessary in order to defeat inflation. Over the next two years the overnight interbank interest rate (the federal funds rate) increased 800 basis points and the 10-year Treasury bond rate increased 600 basis points. Long-term real interest rate differentials between the United States and its trading partners increased 4 percentage points between September 1979 and the end of 1982, contributing to an appreciation of more than 21 percent in the Federal Reserve's real trade-weighted value of the dollar.

Although the Volcker disinflation succeeded by the end of 1982 in breaking the back of inflation, interest rates in the United States remained stubbornly high. The real interest rate differential between the United States and its trading partners increased another 100 basis points in 1983 84 as compared with 1981 82. The real exchange rate also appreciated another 20 percent over this period (Frankel 1994). The most likely explanation for the continued increase in U.S. interest rates and the dollar over this period was the large increase in the U.S. structural budget deficit.

The dollar continued appreciating between June 1984 and February 1985 even though real interest rate differentials, trade imbalances, and other fundamental factors were moving against the dollar

(Boughton 2001). Many argued that the appreciation at this time was due to a speculative bubble. The Reagan administration, however, maintained a position of benign neglect toward the exchange rate. The strong dollar made U.S. exports, measured in foreign currencies, more expensive and U.S. imports measured in dollars cheaper. It thus led to a burgeoning trade deficit. Between 1980 and 1985, the trade deficit increased by $87 billion, reaching $112 billion in 1984. The automobile, steel, textile, agriculture, and capital goods sectors were particularly hard hit. Firms responded to these pressures by seeking tariffs and other protectionist restrictions on free trade. In 1985, for example, more than 300 trade bills were introduced in Congress; of these, 99 were directly and seriously protectionist and 77 were potentially so (Destler 1986).

Coordinated Action to Correct the Global Imbalances These protectionist pressures jolted the Reagan administration and foreign leaders into action. They saw the world trading system that had been carefully built up over 40 years in danger of collapsing. Foreign leaders were also concerned because the strong dollar was raising import prices and generating inflationary pressures.

In January 1985, the incoming treasury secretary James Baker suggested that the Reagan administration should reconsider its policy of benign neglect toward the exchange rate. In February 1985, the G5 countries engaged in coordinated interventions in foreign exchange markets to lower the value of the dollar. The dollar began falling in February 1985 and fell 15 percent against the German mark and 10 percent against the Japanese yen before the signing of the Plaza Accord in September of that year. It is unclear whether these currency realignments were due to market forces correcting the overvaluation of the dollar (Feldstein 1988) or to foreign exchange intervention by G5 countries in February (Frankel 1994; Funabashi 1989).

The actual agreement sought to correct trade imbalances not only through exchange rate changes but also through expenditure-reducing policies in the United States and expenditure-increasing policies in the other countries. The agreement stipulated:

1. The U.S. would reduce its budget deficit by more than 1 percent of gross domestic product (GDP) in fiscal year 1986 and reduce it further in the future;
2. Japan would liberalize financial markets to ease up on consumer credit and would take into account the yen exchange rate when conducting monetary policy;
3. Germany would cut taxes;
4. All the countries would cooperate closely in foreign exchange markets to encourage further depreciation of the dollar; and
5. All trading partners would resist protectionist pressures (Meyer et al. 2002).

The U.S. current account balance improved over the next several years and returned to balance in 1991. Although several factors, including a recession in the United States, were responsible for this improvement, the orderly exchange rate adjustments following the Plaza Accord were important factors as well. In the first decade of the 21st century, as U.S. trade deficits as a percentage of GDP have reached levels twice as large as the imbalances of the 1980s, there have been calls for a new Plaza Accord (see, e.g., Cline 2005). Since the United States is running trade deficits with many countries, a new agreement to reduce the value of the dollar would probably involve many more countries than the five that took part in the original accord.

See also balance of payments; Bonn Summit; effective exchange rate; expenditure changing and expenditure switching; Federal Reserve Board; foreign exchange intervention; global imbalances; international policy coordination; Louvre Accord; mercantilism; real exchange rate; twin deficits

FURTHER READING

Bergsten, Fred, Olivier Davanne, and Pierre Jacquet. 1999. "The Case for Joint Management of Exchange Rate Flexibility." Institute for International Economics Working Paper 99 9. Washington, DC: Institute for International Economics. Argues that there is a need for more economic policy coordination among G7 countries.

Boughton, James. 2001. *Silent Revolution: The International Monetary Fund, 1979 1989*. Washington, DC: International Monetary Fund. A history of the IMF between 1979 and 1989 by its own official historian.

Cline, William. 2005. "The Case for a New Plaza Agreement." Institute for International Economics Policy Brief 05 4. Washington, DC: Institute for International Economics. Argues that a new Plaza Agreement is needed to help correct global imbalances in the 21st century.

Destler, I. M. 1986. *American Trade Politics: System under Stress*. Washington, DC: Institute for International Economics. A useful history of trade policy in the 1980s.

Feldstein, Martin. 1988. "Distinguished Lecture on Economics in Government: Thinking about International Economic Coordination." *Journal of Economic Perspectives* 2: 3 13. A negative assessment of international economic coordination by an influential economist.

Frankel, Jeffrey. 1994. "The Making of Exchange Rate Policy in the 1980s." In *American Economic Policy in the 1980s*, edited by Martin Feldstein. Chicago: University of Chicago Press, 293 341. A readable and insightful history of exchange rate policy in the 1980s.

Funabashi, Yoichi. 1989. *Managing the Dollar: From the Plaza to the Louvre*. Washington, DC: Institute for International Economics. Provides a thorough, behind the scenes account of how the G5 countries were able to agree on the Plaza Accord.

Meyer, Laurence, Brian Doyle, Joseph Gagnon, and Dale Henderson. 2002. "International Coordination of Macroeconomic Policies: Still Alive in the New Millennium?" International Finance Discussion Paper 723. Washington DC: Federal Reserve Board. A review of the theory of international macroeconomic policy coordination combined with a summary of recent experience in this area.

WILLEM THORBECKE

■ **political economy of policy reform**

Policy reforms create winners and losers. Hence, implementation of policy reforms requires attention to political economy. Even when reforms are efficient in the aggregate, the absence of compensation means that reforms create losers who may oppose reforms. Thus many reforms that may be efficient will not be implemented. Even reforms that will benefit a broad

swath of society are often delayed for long periods of time. The political economy of policy reform studies the implementation of reforms and why they are often delayed.

The political economy of reform can be viewed as an issue of how to get reforms enacted, and then how to keep them from being reversed. The first problem involves relaxing ex ante political constraints. The latter problem is to relax the ex post political constraints (Roland 2002).

An important aspect of implementation is the design of reforms. How reforms are designed may affect both the likelihood of implementation and the sustainability of reforms. The major distinction in this area is between "big bang" and gradual reforms. Big bang reforms are attempts at comprehensive reforms, the idea being that unless a slate of reforms is undertaken simultaneously the policy will be unaffected. Gradual reform focuses on sequencing and the benefits of trial and error in policymaking. The argument is that reform on too broad a front is too difficult to implement. Getting the sequence of reforms right can build momentum for reform; thus the argument is also one of political acceptability. This controversy has been most discussed in the literature on economic transition, though it has more general application. Often Russia is taken (incorrectly) to be an example of big bang reforms and China is taken as the premier example of gradual reforms.

To fix ideas, consider a policy of tariff liberalization in a small economy. Implementation of the reform will raise national income. But it will cause losses among factors in the import-competing sector while leading to gains in those sectors that consume the imports and in the export sector. Assuming the gains exceed the losses the potential for compensation could lead to a Pareto-improvement and thus no problem with implementing the policy. But as compensation is problematic, the potential losers may oppose the policy or at least cause significant delay. This is the problem that is the focus of the political economy of policy reform.

Implementing Reforms Fernandez and Rodrik (1991) developed a model of status quo bias in re-

form that has become classic in the literature. The idea is that when individuals are uncertain over the benefits of reform potentially beneficial programs may not be implemented. Even if everyone knows that in the aggregate the policy will be beneficial (i.e., there is no aggregate uncertainty), if individuals are uncertain over their own situations a status quo bias may still arise. Individual specific uncertainty implies that agents do not know whether or not they will win from reforms. Even if they know that a majority will benefit they may vote against reforms if they believe that they may be losers.

One may suspect that compensation packages may be useful in alleviating this problem, but there is a credibility problem at work. Ex post a majority of voters benefit; hence the losers have little power to enforce a promise for compensation. And with a majority in favor, there will be little chance to reverse the reform. Hence, potential losers know that a promise of compensation is not credible, and this may cause them to oppose reforms. Moreover in practice, it is very difficult to identify who the losers are. This demonstrates that the potential losers have a strong incentive to prevent implementation   their bargaining power drops dramatically after the policy is passed.

The Fernandez-Rodrik model explains why reform may not be implemented, but reforms do take place. To explain why suboptimal policies are reversed requires a dynamic analysis. Alesina and Drazen (1991) developed a model in which interest groups differ over the net benefits of reform. They consider a reform that has public good aspects   a fiscal reform, for example. The problem is that there is a distributional conflict over who wins most from the reform. Each group prefers that the other group pay a higher share of the costs. Delay in reform is modeled as a war of attrition. Reform occurs when one of the groups realizes that it has more to lose from delay than from reform. When one side capitulates, the reform occurs. This also implies that reform will take place sooner the greater the crisis, since this makes delay more problematic. The problem of reform is then to overcome distributional concerns to create consensus over reform.

While the Fernandez-Rodrik model emphasizes winners and losers (and is applicable to reforms that need not have public good aspects) the Alesina-Drazen model emphasizes the need to form consensus for reform to eventually occur. What is common to both, however, is the emphasis on heterogeneity and the conflict of interests that this implies. The models are essentially complementary.

Another mechanism that may lead to blockage of beneficial reforms reflects the influence of special interest groups. Grossman and Helpman (1994) analyzed how such interest can have influence on political decision making (through monetary contributions or information transmission) and used their model to explain the structure of trade protection. Politicians care about social welfare and campaign contributions that can be used to get votes in future elections. They show how even if politicians care about social welfare, the campaign contributions that can be raised by running on a platform of trade protection, for instance, may cause them to place greater weight on workers and capital owners in some sectors of the economy than on others (Grossman and Helpman 1994).

Big Bang versus Gradualism An important debate in the literature on policy reform concerns the *dimensionality* of reform: Shall reform be piecemeal or comprehensive? In the literature on economic transition, the debate has centered on two dimensions: speed and comprehensiveness. The form this debate has taken is often characterized as shock therapy (or big bang) versus gradualism. Related is the issue of whether there is a single blueprint sometimes called the Washington consensus that suits all reforming countries. It is important to note that this debate often conflates two dimensions of reform: timing and scope. This is perhaps inevitable, as big bang reforms typically push for rapid, comprehensive reform. But it must be noted that gradual reforms can be comprehensive as well, but over time.

Big bang reform packages are obviously comprehensive in intent. It is crucial to emphasize, however, that gradual reform is not at all the same thing as partial reform. The primary debate is over pace and sequencing. In the case of emerging market reforms,

the debate centers on whether financial institutions and the rule of law must be implemented before capital markets are liberalized. This argument has been made most forcefully by McKinnon (1993), for example. The broader issue is whether markets can operate before the institutions that support them are established, or whether broad-brush market reform creates an environment where the institutions will develop.

Especially in the transition context, the goal is eliminating the command principle and moving to a private property based market economy. Hence, *eventually* the same steps must be taken. The question is how many steps to take at once and whether they must be taken in a precise order or not. For example, do you need to establish institutions of a private property economy before privatization, or will privatization create the *demand* for private property institutions? There is little debate over the importance of institutions for the success of the market economy, but there is still debate over how to induce their development.

The competing principles are easy to understand. The need for a big bang seems almost implied by the lessons from partial reform (e.g., Murphy, Shleifer, and Vishny 1992). Partial reforms for example, enterprise autonomy without price liberalization lead to supply diversion and inferior outcomes. If reforms are complementary, then partial implementation may lead to outcomes inferior to the status quo. When reforms are complementary, implementation of the whole package may be necessary for the reforms to succeed. Moreover partial reforms in this context create rents for insiders who then may oppose further reforms. And if there is a window of opportunity for reform, it may be advantageous to implement a far-reaching reform before the window is shut.

For example, in the case of transition there is a question of how much of the infrastructure of a market economy is needed for markets to be effective. Privatization cannot work without price liberalization, and neither policy can work without stabilization and the elimination of the soft budget constraint. If the command system is not dismantled the private sector will not be able to compete. The big

bang view thus argues for reform on a broad front so that complementarities across reforms can be utilized.

The case for a big bang approach often builds on the argument that it takes some time for people to realize whether they are going to win or lose from reform, and to form a powerful opposition against reform if they expect to lose, so by moving quickly, the politician prevents emergence of opposition powerful enough to actually stop or reverse reforms. This is often referred to as the "window of opportunity." It was a prime consideration of reformers in Poland and the architects of privatization in Russia.

A related argument for the big bang stems from the problem of state capture (Hellman 1998; Sonin 2008). Suppose that a complete set of reforms must be implemented and that reforms imply short-term costs. Then *if* the reform package gets stuck partway, voters may fear they will reap only costs and no benefits and hence they will oppose reforms all together. This may happen when the reform process is likely to be captured by winners from partial reform. Hence when the political environment is such that this may occur, big bang policies may be the only way to implement reforms. Forward-looking agents that would be willing to bear the short-term costs if they were sure to reap the long-run benefits may withdraw their support for reform if the realization of the long-run benefits becomes uncertain because of the risk of getting stuck in a partial reform equilibrium.

Another potential argument for the big bang approach (noted by Olofsgård 2003) can be derived from the recent literature on self-control problems. The idea is that voters may not have the self-control to stick with a schedule of partial reforms. The urge for immediate gratification may induce voters to delay reforms or to support the nonreformer in order to avoid the short-term cost of reforms. This may occur even though voters' lifetime expected utility as calculated *in the very beginning of the reform process* is higher with continuation of reform. The big bang approach may thus serve as a commitment device in this case, because it restricts the choice set of the voters if reform reversal is unlikely or very costly.

The alternative view — dubbed the "evolutionary-institutional view" by Roland (2002) — is motivated,

first of all, by the recognition that one cannot do all at once. There are limited skills and limited amounts of time available to policymakers. This alternative conception gradualism sees the trial-and-error method as a virtue. Gradualists often accuse proponents of the big bang approach of hubris. McMillan (2004, 35), for example, argues: "Reform is hard to do because we cannot predict its effects. The big bang approach presumes we know where we are going and we know how to get there. But we don't." This argument, also put forward by Easterly (2006) and Rodrik (2008), emphasizes humility and a menu of approaches rather than a unique program applicable in all countries.

This argument is extended to the issue of complementarities in reform. Although the necessity for implementing a package of reforms would seem to argue against gradualism, McMillan (2004, 36) argues to the contrary: "The systemic interactions are hard to predict. We know little about how the pieces of the market system fit together. Some interactions may be complex and indirect, and so we may not even be able to anticipate their existence. Others are straightforward . . . but we have little data on their magnitude."

Opponents of big bang reforms point out that if they are designed incorrectly they can lead to a lock-in of bad reforms. Sonin (2003) argues that mass privatization in Russia created a lock-in of power among early winners. These oligarchs could afford their own property rights protection and thus were opposed to the development of the rule of law, government reform, tax reform, and other policies that would threaten their own incumbent position. Indeed this argument has been made in a more general way by Rajan and Zingales (2004), who argue that the biggest threat to capitalist economies often comes from incumbents. It is not clear, however, to what extent this is an argument against big bang so much as against bad sequencing of reform. It is certainly the case that bad reform will lead to bad effects.

The classic example pointed to by proponents of gradualism is the dual track system in China. This system liberalized at the margins: enterprises were

required to continue fulfilling the central plan but could use the market to sell above-plan-level production. Li (1999) provides an interesting analysis contrasting the big bang and dual track reforms with special emphasis on the Chinese experience. The success of Chinese reforms is frequently attributed to the dual track. This makes it all the more ironic that the dual track closely resembles Soviet partial reforms   the "law of state enterprises" of 1987   which performed so poorly. This suggests that the success of reforms may be attributable to the specific conditions of the case rather than to general principles. What worked in China did not work in the Soviet case, most likely because state authority remained strong in China but was deteriorating rapidly when these reforms were implemented in the Soviet Union.

To some extent, gradualism may require that political control, à la China, remains. This seems to be the lesson of the comparison of Gorbachev's reforms regarding the Law on State Enterprises with the dual track system in China. Both allowed enterprises to use markets to allocate above-plan-levels of production. In China the system has been a success, as enterprises continued to produce for the plan, although significant corruption occurred as well. In the Soviet Union, on the other hand, plan fulfillment fell dramatically as supplies were diverted to take advantage of higher market prices, and as a result, planners increased targets continually. The lack of enforcement power of the government made it hard to utilize this mechanism in the Soviet Union. Without such control it may be difficult to control the pace of transition. And democracy may work against more gradual mechanisms. This suggests that often the difference between the relative successes of the two approaches may be due to a third, independent, factor: the nature of the political equilibrium in a reforming country.

A related argument in favor of gradualism stems from the notion of an incubation period. In the late Soviet period there was the idea of the "500-days" plan. This was a blueprint for reform, written by leading reformers and accepted by Yeltsin that was eventually rejected by Gorbachev. The essence of the plan was to move in a steady way to a market economy. Enterprises would be broken up before being privatized to reduce monopoly power. Apartments would be privatized before price liberalization to soak up the monetary overhang. The key idea was to prepare the system for the transformation.

The opposite view is that crisis prevents an incubation period. This is certainly true in the Polish case and also in the Russian case. The problem is that the urge to undertake radical reform did not take hold until a crisis emerged, at which point there was no time left for incubation. This argument is less important in the case of Czechoslovakia and Hungary, where in fact some incubation did take place. These considerations point to the fact that the amount of hemorrhage in the fiscal budget may be a critical factor in deciding the outcome of reform.

Sequencing is also important for the continuation of reform. If early reforms lead to an improved situation for the median voter, this may create support for future reforms. The success of agricultural reform in China is often seen as creating a favorable environment for reforms of industry. This suggests that starting with a reform that increases the welfare of a majority runs a smaller risk of reversal. In contrast, incorrect sequencing (starting transition with the more painful reforms) undermines popular support and may unnecessarily lead to reform reversal. The case for gradualism thus crucially hinges on correct reform sequencing (Dewatripont and Roland 1995, 1209).

One way to cope with the status quo bias is to provide compensation. This need may also argue for partial reform as it reduces the fiscal cost of compensation. For example, unemployment is very likely to increase during transition. To avoid opposition to a reform by the unemployed, the government needs to provide unemployment benefits. In these conditions, Dewatripont and Roland (1995) point out that a slower, feasible speed of transition may be preferable because the pressure on the fiscal sector is lessened. Of course fiscal costs are not the only costs that should be considered, but imperfect reforms that are actually implemented may be preferable to perfect reforms that cannot be implemented.

The importance of political constraints may thus alter the balance between big bang and gradual reforms.

Reversal Costs A big bang strategy involves high reversal costs, which are often considered to be an advantage ex post since they reduce the reversibility of enacted reforms, which is a constant concern for reformers. Thus Anatoly Chubais, the architect of Russian privatization, spoke of the irreversibility of mass privatization as one of its most important benefits. If reforms are irreversible, then a concerted effort can achieve a permanent impact on performance. The problem with this view, however, is that it ignores the ex ante problem. If reversal costs are high, then it may be harder to implement reforms in the first place. After all, if reversal costs are high and reforms have unexpected consequences, any learning cannot be acted on. Hence if there is significant uncertainty about the consequences of reform that is, if there is significant probability of learning then high reversal costs make it harder to enact reforms in the first place. From the ex ante point of view, however, high reversal costs in the case of a negative aggregate outcome may make a big bang approach politically unfeasible. Gradualism makes reforms easier to start because it gives an additional option of early reversal at a lower cost after some learning has taken place (Dewatripont and Roland 1995).

If partial reform is less costly to reverse than full reform, gradual reforms may be more politically acceptable than full reforms because they provide an option of early reversal (Dewatripont and Roland 1995). If after partial reform is implemented, a continuation of reform toward full reform seems unattractive to a majority because the signals given by partial reform about the future are not promising enough, then it is always possible to come back to the status quo. On the other hand, if the signals given by early reform are promising enough, then the reforms can continue with greater support. Gradualism thus lowers the cost of experimenting with reform and makes a move away from the status quo more acceptable to a majority.

Note that this argument for partial reform is intimately connected with learning. If reforms are complementary, then one may learn from partial reform. Suppose that one can decompose a reform package into two parts. A gradual reform would implement one part first price liberalization in agriculture before industry, for example. One may then learn about the consequences of the package from experiencing the first part. Informativeness is the key necessary condition for gradualism to dominate the big bang strategy in the sense that learning about the first reform tells whether to try the second reform or not, depending on the realization of the state after the first reform is implemented. If there is no complementarity, there is nothing to learn and hence you might as well do all at once.

Costs and Benefits of Gradual Reform We can summarize this discussion by considering the costs and benefits of gradual reform. The most important cost is simply the delay in improved performance. If some important reforms are only implemented later, then the full benefits of reform are delayed. An equally important cost may be that delayed reforms continue the hemorrhaging that is taking place. This is especially true when reforms are undertaken in economies suffering crisis. When the IMF is called to provide support for an economy it is already in an economic crisis. Similarly, in the case of transition economies, the status quo may not be feasible. The economy is in crisis and gradual reform may not be possible. A third important cost of gradual reform is that it may dissipate the credibility of reformers, given that partial reform was a frequent failed policy in planned economies.

The benefits of gradual reform are also straightforward. Most important, perhaps, is that if it is carried out successfully, momentum for further reforms can be built. Building constituencies that support reform may enhance the sustainability of reform. Gradual reforms also allow for learning and may be easier to implement when reversal costs are high. They allow for experimentation. Finally, gradual reforms may involve a lower level of transfers, and hence, a more feasible fiscal burden.

On balance, the choice between big bang and gradual reforms may come down to the extent of current crisis and the strength of the state. If there

is significant political legitimacy or control, gradual reforms may be feasible. If reforms are taking place in a period of extraordinary politics, however, big bang reforms may be the only choice.

Empirical Evidence Given the experiences of emerging and transition economies what has been the experience of big bang versus gradualism? The question is not easy to answer, though advocates of each approach have plenty of cases to point to. The experience of Russia and China are often contrasted and used to illustrate the relative success of big bang and gradual reform policies, respectively. The problem, however, is that policies are not implemented in isolation. The success of policies will depend on external circumstances. Russian reforms took place during a period of very low oil prices. When the economy deteriorated, it was natural to blame reforms. But this may miss the essential point. For an economy that is the second largest exporter of oil and whose economy is heavily resource-dependent, this fact accounts for economic performance far more than any reform packages.

A related problem is that the choices between the two programs are not chosen randomly. The choice is endogenous, and the political equilibrium in the country is often decisive. In more stable societies gradual reform may be chosen, but if the society has a stable political system even big bang reforms may be successful. In contrast, most economies in crisis choose big bang approaches, and performance is frequently less than stellar. But there is a reason these economies were in crisis, and it is not clear that a gradual program would succeed in these countries.

Another fundamental problem with empirical work focusing on this question is how to identify whether a country chose a big bang or gradual approach. Often, this judgment is made ex post via expert surveys. The identification is thus influenced by how successful reform was thought to be. But this just ensures that one's prior assumptions are confirmed. If a country announces a big bang reform but is only partially successful at implementing the various parts of the program, should this country be classified as big bang or gradual? This is a knotty problem that has plagued this literature and is a major reason for a lack of consensus about outcomes.

See also evolution of development thinking; international institutional transfer; political economy of trade policy; structural adjustment; transition economies; Washington consensus

FURTHER READING

Alesina, Alberto, and Allan Drazen. 1991. "Why Are Stabilizations Delayed?" *American Economic Review* 81 (5): 1170 88. Discusses why stabilization packages are typically implemented only after long delays.

Dewatripont, Mathias, and Gerard Roland. 1995. "The Design of Reform Packages under Uncertainty." *American Economic Review* 85 (5): 1207 23. Examines how uncertainty affects the design of reform packages.

Easterly, William. 2006. *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good*. New York: Penguin Press, 2006. An account of why foreign aid is ineffective, focusing on the lack of accountability.

Fernandez, Raquel, and Dani Rodrik. 1991. "Resistance to Reform: Status Quo Bias in the Presence of Individual Specific Uncertainty." *American Economic Review* 81 (5): 1146 55. Demonstrates how the presence of uncertainty regarding the distribution of gains and losses from reform can prevent efficiency enhancing reforms from taking place.

Hellman, Joel. 1998. "Winners Take All: The Politics of Partial Reform in Postcommunist Transitions." *World Politics* 50 (2): 202 34. Argues that the process of economic reform in transition is threatened less by the losers than by the net winners.

Grossman, Gene M., and Elhanan Helpman. 1994. "Protection for Sale." *American Economic Review* 84 (4) (September): 833 50. Develops a model where special interest groups make political contributions in order to influence an incumbent government's choice of trade policy.

Li, Wei. 1999. "A Tale of Two Reforms." *Rand Journal of Economics* 30 (1): 120 36. Identifies conditions under which a big bang reform reduces output initially while a Chinese style reform increases output.

McKinnon, Ronald. 1993. *The Order of Economic Liberalization: Financial Control in the Transition to a Market*

Economy. 2d ed. Baltimore: Johns Hopkins Press. Out lines the appropriate sequencing of liberalizing govern ment policies in domestic finance and foreign trade for securing open markets.

McMillan, John. 2004. "Avoid Hubris and Other Lessons for Reformers." *Finance and Development* 41 (3): 34 37. Argues for a gradual reform process and against big bang reforms.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny. 1992. "The Transition to a Market Economy: Pitfalls of Partial Reform." *Quarterly Journal of Economics* 107 (3) (August): 889 906. Demonstrates how partial reforms in a planned economy can cause performance to dete riorate via diversion of inputs.

Olofsgård, Anders. 2003. "The Political Economy of Reform: Institutional Change as a Tool for Political Credibility." Mimeo. World Bank. December 4. Argues that institutional change is a critical step for reformers to obtain credibility.

Rajan, Raghuram G., and Luigi Zingales. 2004. *Saving Capitalism from the Capitalists*. Princeton, NJ: Princeton University Press. Asserts that capitalism's biggest threat is that insiders continually try to eliminate competition and that financial markets are the best means to preserve open markets.

Rodrik, Dani. 2008. *One Economics, Many Recipes: Globa lization, Institutions, and Economic Growth*. Princeton, NJ: Princeton University Press. Argues that globaliza tion reduces poverty only when countries tailor policies to their own specific constraints.

Roland, Gerard. 2002. "The Political Economy of Transi tion." *Journal of Economic Perspectives* 16 (1): 29 50. Argues that the experience of transition economies demonstrates the superiority of the evolutionary insti tutional approach.

Sonin, Konstantin. 2003. "Why the Rich May Favor Poor Protection of Property Right." *Journal of Comparative Economics* 31 (5): 715 31. Shows that in unequal soci eties the rich may benefit from shaping economic insti tutions in their favor, in particular by weakening public property rights enforcement.

———. 2008. "State Capture in Transition: Corruption." In *New Palgrave Dictionary of Economics*. London: Palgrave. Analyzes how during transition the fear of the *Leviathan* state has been transformed into the reality of

the captured state, where powerful large interests con trol the playing field.

BARRY W. ICKES

■ **political economy of trade policy**

After the provision of internal order and external security (and bound up with both as a source of revenue), the regulation of international trade is one of the oldest known activities undertaken by states. Thus it is not surprising that we have records of an alytical comment on commercial policy from the earliest writings on economics and political econ omy. Most of this early writing is "political econ omy" in the sense of practical policy advice, but this is not unrelated to the term's more modern meaning as positive accounts of policy choice. Policy advice is given under an implicit model of government (sometimes called the Weberian model after Max Weber) that the government is willing and able to act on advice intended to improve aggregate social wel fare. Given its focus on explaining welfare-worsening policy, contemporary political economy research generally rejects this model (even when, as in the case of very low levels of protection, it would seem to be a more accurate model than the obvious alternatives). Instead, we take the obvious fact that trade policy deviates from virtually any straightforward model of welfare maximization as a warrant to consider alter native accounts of policy determination. The great majority of this work begins with an attempt to link policy preferences of the relevant agents to their material self-interests, defined in terms of the prim itives (axiomatic assumptions) of some underly ing model of the economy, and then seeks to map these policy preferences into policy outcomes via some (usually extremely spare) model of policy determination.

Early Steps in the Political-Economy Analysis of Trade Policy The most immediate predecessors of modern research on the political economy of trade policy are three works by political scientists that still repay close study and that contain most of the themes that still motivate contemporary research:

Schattschneider's *Politics, Pressure, and the Tariff* (1935); Bauer, Pool, and Dexter's *American Business and Public Policy* (1963); and Lowi's "American Business, Public Policy, Case Studies, and Political Theory" (1964). The first of these treats trade policy (in this case the Hawley-Smoot tariff) as the outcome of an asymmetric lobbying process in which protection seekers dominate and certain sectors are more effective than others. Bauer, Pool and Dexter saw their work as a rejection of Schattschneider's. Looking at omnibus trade legislation in the period 1953 62, they found (among other things) members of Congress to be relatively free from lobbying constraint and saw lobbying primarily as information transmission. Lowi's brilliant review and reconciliation argued that a fundamental institutional change had resulted in a redefinition of trade policy such that the politics had changed between the periods studied in these two classic works. Although the great bulk of contemporary research on the political economy of trade is rooted in the asymmetric group theoretic tradition presented in Lowi, a growing body of work emphasizes institutional features and, while not much applied to trade policy, the dominant strand of research on lobbying in political science now emphasizes informational issues.

The first wave of research by economists on the political economy of trade protection was actually less concerned with developing a coherent and, at least in principle, testable positive theory than with finding a way to increase the costs of protection. In the first paper of this sort, Tullock (1967) argued that the standard welfare analyses (focused on what economists call "deadweight loss" or "welfare triangles") were sizable underestimates of the actual welfare costs of protection (and monopolies). Specifically, potentially productive resources were spent seeking the rents implied by the presence of the tariff. Accounting for these would permit us to count the rent rectangle, as well as the deadweight loss triangles, as costs. The empirical application of this logic by Krueger (1974) to the case of Turkey, in which rent seeking clearly added to the costs of protection, moved political economy analysis to the center of trade policy analysis, at least in developing countries.

This approach reached its most sophisticated form in the work of Bhagwati and his coauthors on directly unproductive profit-seeking activities (e.g., Bhagwati and Srinivasan 1980). As applied to pathological situations (e.g., Turkey) this approach makes a lot of sense, but as a general extension of the normative analysis of trade policy, it is deeply problematic: adding the economic costs of political action to the costs of protection without a systematic normative analysis of political action makes little sense. After all, in a nonpathological situation, lobbying or directly unproductive political activities are just other terms for citizens approaching their government. The ability of citizens to do this is an essential component of the democratic political process, something most of us would consider highly positive.

The other precursor to current work grows out of Olson's *Logic of Collective Action* (1965) and the closely related earlier work by political scientists on the foundations of group theoretic models of political action. Olson's key insight, relative to much of the earlier research on the political activity of groups in the determination of policy outcomes (called "pluralist theory" by political scientists), was that the decision to invest real resources (e.g., time and effort, as well as money) in group membership was a rational decision, constrained by the expected impact of those resources on the goals of the group. Because groups, especially political groups, seek goals that benefit all members, regardless of their level of investment, the decision to invest has the form of a decision to provide a public good (i.e., it has a prisoners' dilemma structure). In addition, asymmetries in group effort can be expected to emerge based on group size. That is, large groups will find it hard to provide the optimal level of the public good, while small groups with concentrated benefits from political action will find it relatively easier to organize. Thus small groups expecting concentrated benefits can be expected to be more politically successful than large groups with diffuse benefits.

In the case of trade policy, the most common form of this argument is that producers of an import-competing good will expect concentrated benefits (not only supporting higher factor returns, but

protecting jobs and immobile capital from unemployment) and will find it in their interest to organize to support protection, while a large number of consumers, each of whom bears only a small additional cost as a result of the protection, will not find it in their interest to organize. Thus we will observe protection even if it is welfare-reducing in aggregate. Although this may well explain why we do not observe much in the way of consumer activity on trade policy, it does not explain why large retail intermediaries (e.g., Wal-Mart) or industrial users of intermediate goods are not more active on trade, or why what little consumer activism we do observe tends to support protection (e.g., Public Citizen, a U.S. consumer rights organization founded by Ralph Nader).

A sizable empirical literature developed to account for tariff structure in terms of a number of more or less ad hoc variables in a linear regression framework. Sign patterns in these variables were suggested in terms of a number of loose explanatory frameworks. This work increasingly focused on variables seen to be associated with effective group formation and then used loosely construed models linked to early Chicago school political economy as an interpretive framework. Overall this work provides quite strong evidence that political-economic factors must be part of any coherent account of systematic patterns in the tariff structure.

Political Preference and Minimally Institutional Political Economy The core of virtually all political economy models developed by economists is a formal derivation of preferences over policy from primitives of the underlying model of the economy. More formally, endogenous policy models build some form of explicit political structure into some form of neoclassical general equilibrium model. That is, the underlying economy is made up of households, characterized by preferences over final goods and portfolios of productive factors, and firms, which transform factors of production into final goods. We can denote a neoclassical economy with the set: $E = \{\mathbf{Z}, \mathbf{F}, \mathbf{R}\}$, where $\mathbf{Z}$ is a matrix allocating factors of production among households, $\mathbf{F}$ a vector of technologies, and $\mathbf{R}$ a vector of household prefer-

ences over final commodities. To this economy we attach a vector of possible interventions ($\mathbf{t}$) and a political mechanism ($\mathbf{M}$) yielding a political economy: $\Pi = \{\mathbf{Z}, \mathbf{F}, \mathbf{R}; \mathbf{t}, \mathbf{M}\}$. The easiest part of the analysis of a system like Π involves the derivation of citizen preferences over policy. For a given household, $h(R_h \in \mathbf{R}, \mathbf{z}_h \in \mathbf{Z})$, for a fixed E, we simply ask how any relevant state of the policy variable (t) affects household welfare. By answering this question for every feasible state of the policy variable, we trace out household political preferences over that policy.

It is now well known that the most general version of this model (usually called the Arrow-Debreu-McKenzie model) does not yield comparative static results of the sort necessary for this exercise. As a result, most work in this area builds on either a Heckscher-Ohlin-Samuelson (HOS) model (i.e., a model in which all factors are fully mobile between sectors) or a Ricardo-Viner model (i.e., a model in which every sector is characterized by a specific-factor and a single mobile factor is used by all sectors). The former yields conflict based on factor ownership (i.e., class conflict), while the latter yields conflict based on industry attachment. Recently, a specialized form of the latter has been widely used, in which n-1 sectors are characterized by a specific factors structure and the nth sector (the *numeraire* sector) is a freely traded Ricardian good (Grossman and Helpman 1994). This is usually combined with quasilinear preferences to produce an economy in which virtually all general equilibrium linkages have been severed. This makes for an extremely tractable, if somewhat implausible, framework for political economy modeling of contemporary trade policy. If we think of the HOS model (and its dimensional generalizations) as a model of the long run and the Ricardo-Viner model as representing a relatively short run, it is probably clearest to treat their use in political economy models as reflecting the period of political calculation. That is, if political agents make their calculations over a relatively short time period, the most appropriate model would be some form of specific-factors model; while, if the time horizon of political calculation is long, the HOS model is more appropriate.

The evidence is somewhat mixed, but the bulk of it would seem to suggest that, for most agents, the time horizon of political calculation is relatively short. The most compelling evidence comes from actual political activity. Starting with Magee's (1978) classic study of testimony on trade legislation, most work has found strong evidence of sector-based political activity. This is consistent with studies of capital and labor mobility that suggest relative immobility over politically plausible time horizons (i.e., two to six years in the United States). Research on policy preferences revealed by public opinion polls is generally interpreted as more consistent with factor-based calculation, but the implications of these results for either factor mobility or the time horizon of political calculation are currently a matter of considerable dispute.

Just as knowing preferences over final goods is not sufficient for a theory of market equilibrium, knowledge of political preferences is not sufficient to determine political-economic equilibrium. In principle, a model of political action must be combined with a model of policy determination to determine a full political-economic equilibrium. Here there are many possibilities. One convenient way of distinguishing among the basic models is in terms of what is assumed about the activity permitted of demanders of policy and suppliers of policy. If we consider that each type of agent (demander and supplier) may be active or passive, we get the simple typology shown in table 1.

The simplest approach is to assume that policy is determined by a referendum on the tariff. Under the

assumption that voting is costless (and the assumption that preferences are single-peaked over a one-dimensional t), this approach simply determines the policy outcome at the most preferred point of the median voter. No resources are used up in the political process, there are no gains from misrepresentation of preferences, so once citizen policy preferences have been determined, the step to final policy determination is straightforward (Mayer 1984). The virtue of this approach is its simplicity, but it has at least two major drawbacks as a framework for empirical work: there have been almost no actual referenda on trade policy, and at the level for which general equilibrium analysis is appropriate, trade policy is inherently multidimensional. As a public issue, trade policy usually is discussed in terms of the overall level of protection offered in the country as a whole (not in terms of tariffs on specific line items in a tariff schedule). Seen in this way, there have been times and places in which "The Tariff" was sufficiently important to public partisan competition that a given election could be seen as a referendum on the tariff (e.g., the Canadian election of 1911 and the British elections of 1906 and 1914).

An extension of this approach implicitly assumes that representatives are elected to represent their constituencies *on trade policy*, so that their votes on trade policy can be treated as determined by the material interests of their constituents. However, as Bauer, Pool, and Dexter argued, and as implied by theoretical work on voting in high dimensional issue spaces, given the dimensionality and complexity of the issue environment in which representatives operate, they will generally be free to vote on virtually any issue as they prefer. Thus, especially for issues such as trade, which are not major reelection issues, inference based on this constraint seems less than well grounded. Given that, at least until recently, trade policy has not been a major public issue, some form of lobbying model would seem to be a more solid basis for framing the analysis of the political economy of trade policy.

The simplest approach is to follow Peltzman (1976) in doing away with the attempt to model demanders and focusing on the behavior of suppliers

Table 1
Political economic equilibrium

	Groups are	
	Passive	*Active*
The state is		
Passive	Referendum	"t" formation function
Active	Political response function	Menu auction

who face an untheorized (but completely plausible) function of implicit group demands and a concern with the general welfare consequences of their actions. Hillman (1982) develops a model of this sort applied to trade policy. The *political response function* reflects the trade-off facing a politician who seeks to favor one group (say, producers) at the expense of another (consumers). While Hillman motivates his analysis by loose reference to lobbying in the context of an election constraint, neither the lobbying nor electoral competition are explicitly modeled. The general pattern of signs could emerge from a variety of institutional environments. Thus it seems sensible to characterize this approach as "active state/passive groups." The essentially ad hoc form of the political support function makes for a natural match with the ad hoc empirical analysis referred to in the previous section.

Where the political support function approach abstracts from the political activity of groups, lobbying models, with a passive register state, introduce an explicit analysis of costs of political action and rational strategic behavior by groups, but abstract from active decision making by the state. Findlay and Wellisz (1982) develop a model of this sort for the case of trade policy. The core institutional assumptions of this model are: (1) effective political demand is represented by the lobbying activity of entrepreneurial political action committees (PACs); (2) political activity involves no fixed costs, and there are no collective action problems in the organization of group activity; and (3) the state is a passive register of effective citizen demand. That is, politically active sectors hire labor to engage in lobbying, and the state is represented by a *tariff-formation function*: $t = \tau(\boldsymbol{L})$ where $\boldsymbol{L} = \{L_1, \ldots, L_n\}$ is a vector of quantities of labor hired for lobbying. The outcome is determined as the Nash equilibrium in lobbying, where PACs take into account the direct cost of lobbying (i.e., wL_i), the indirect costs (i.e., the effect on the market wage of withdrawing labor from production for use in lobbying), and the return on lobbying via $\tau(\boldsymbol{L})$.

A major advance in the formal theory of lobbying was made by Grossman and Helpman (1994), who developed a model in which the state and economic agents are politically active, thus rendering both the political support function and the tariff-formation function approaches essentially irrelevant. Grossman and Helpman's essential insight was that the relationship between lobby groups and the state could be effectively modeled as one of common agency (with politically organized sectors as the principals and the state as the common agent) and that the menu auction model of common agency was a tractable framework for developing a broader model of political-economic equilibrium than had previously been presented. Each active PAC offers the state a menu associating a payment with every possible tariff *schedule* and the state then chooses a tariff schedule that maximizes its welfare. Unlike the passive register state, the Grossman-Helpman state is explicitly concerned with overall welfare as well as bribes. Specifically, their state's objective function is a linear combination of bribes and aggregate welfare:

$$G = \sum_{i \in I} C_i(\mathbf{p}) + aW(\mathbf{p})$$

where the C_i are the bribes ("contributions"), W is the aggregate welfare, both conditional on the state's choice of $\mathbf{p}$ (the price vector determined via the small country assumption by the tariff), and a is the weight the government attaches to aggregate welfare. Note how this structure embeds both the Weberian state and the passive register state as special cases.

Given the menu auction structure as a characterization of politics, and the claim that truthful equilibria are focal among the continuum of equilibria that generally result, and the Grossman-Helpman economy characterized by quasilinear preferences, n-1 sectors characterized by a specific factors structure, and a freely traded Ricardian numeraire sector, Grossman and Helpman show (proposition 2) that equilibrium trade taxes and subsidies must satisfy a *modified Ramsey rule*:

$$\frac{t_i^\circ}{1 + t_i^\circ} = \frac{I_i - \alpha_L}{a + \alpha_L} \left(\frac{z_i^\circ}{e_i^\circ} \right),$$

where I_i is an indicator variable that takes a value of 1 if the sector is organized and 0 otherwise, α_L is the fraction of the total population that is organized by some lobby, z_i° is the ratio of domestic output to

imports, and e_i° is the elasticity of import demand. This form has been widely taken to be a framework for structural estimation. Overall, there seems to be considerable econometric support for lobbying as a key determinant of the cross-section pattern of protection.

For all the success of the basic models reviewed in this section, primarily in providing a general account of deviations from welfare optimal policies in terms of political pressure and in providing a specific account of the cross-section pattern of protection, it is hard to escape the conclusion that, more than other branches of research on political economy (in particular, local public finance and macroeconomic policy), the gap between the models and their object (either as formal empirics or as a framework for understanding) is large. The next section considers several avenues of current research that seek to extend the scope of research on the political economy of protection, but we conclude this section with one problem fundamental to lobbying models: exogeneity of group organization.

All of the research on lobbying we have considered to this point treats group organization as exogenous: there is no formal analysis of groups at all in the political response function literature; tariff-formation function and menu auction models treat active groups as fully organized and completely efficient in extracting resources to pursue political goals (the menu auction models explicitly treat organization as primitive reflected in I_i in the modified Ramsey rule). This was a sensible strategy in the early development of lobbying models, but before they can be treated seriously as frameworks for empirical analysis, group formation needs to be systematically integrated into the analysis. The few attempts that have been made to construct such an analysis build, sensibly, on Olson's (1965) theory of collective action. Two approaches seem to have been pursued, both of which treat lobbying as an example of voluntary provision of a collective good. The approach that matches the lobbying models most directly focuses on the decision to join under the assumption that agents joining a PAC will be taxed optimally by the PAC organizer (Mitra 1999). While this is a good

match to the theory of lobbying, it is a poor match to the empirical reality in which groups seem not to be either perfectly organized or perfectly unorganized, but rather to be more-or-less well organized. The alternative is to assume that all groups are potentially active and consider contributions from zero to some maximal (or optimal) level. While there is no shortage of work that considers contributions to a single lobby (treated as a public good), there is very little work that considers those contributions in a strategic context. The problem with this latter approach is that "groups" are not really organized but simply exist as a function of a common pattern of individual contributions. One of the early fundamental criticisms of Olson's work was the lack of attention to political entrepreneurship in the formation of groups, and it seems like an explicit introduction of political entrepreneurs might be a route to linking these two approaches.

Frontiers of Research on the Political Economy of Protection The essentially context-free theoretical environment of the models we have considered to this point is a feature, not a detriment. As with virtually all good modeling programs, the endogenous policy approach to political economy modeling achieves its cutting power by abstracting from details that interfere with the main line of the story in this case, the contest of material self-interest projected from the economy to the political system. Models serve a number of purposes:

1. *Cautionary tales*: most people think X, but in my model X does not occur;
2. *"Just So" stories*: here is a fact, and I can construct a model rationalizing that fact;
3. *Loose frameworks for thinking about issues*; and
4. *Generators of structural frameworks for econometric analysis.*

The Tullock-Krueger-Bhagwati work discussed earlier was essentially of the first sort (protection is more costly than you might think as a result of rent-seeking costs); most of the work discussed after that is of the second and third sort (it is plausible to think about "bad policy" as the outcome of a political process and the empirical work is loosely consistent

with those accounts). Grossman and Helpman's modified Ramsey rule has been treated as a structural form, but the results of that work do little more than underwrite the use of the model for purposes 2 and 3. Making these models more operationally relevant means introducing more of the relevant context into the models. Current work focuses on both political and economic context.

In terms of political context, one approach that is receiving attention is the attempt to explicitly model the link between contributions, elections, and policy outcomes (rather than burying the link in a political response function or the government's objective function). The focus on the role of partisan competition opens the door to both theoretical and empirical work of a comparative nature, as well as cross-sectional work on the foundations of partisan support. This work is interesting, but given that protection is now rarely set by legislatures, it seems unlikely that this line of work will be much help in extending the applicability of current models. By contrast, the sizable body of work that builds on detailed knowledge of the GATT/WTO-legal administered protection mechanisms has already begun to provide a rich body of theoretical and empirical work consistent with the more general models discussed above (see Blonigen and Prusa 2003; and Nelson 2006 for surveys).

An alternative direction of generalization abstracts from the assumption that markets are complete and perfect. One of the standard results of time-series work on the correlates of protection is that demand for protection, and protectionist outcomes, are positively related to unemployment. This is loosely consistent with public opinion data suggesting that support for protection increases if questions are "framed" in terms of unemployment. This suggests that modeling the economy, on which the political economy is based, as characterized by equilibrium involuntary unemployment, could make a major contribution to our understanding of the political economy of trade policy. Recent work by Bradford (2006) and Matschke and Sherlund (2006) make a start in this direction, but there is room for much more. Similarly, standard models assume that po-

litical preferences are strictly self-regarding. Experimental results and public opinion data suggest, however, that political preferences are more complex. In particular, it seems clear that widely held notions of fairness strongly affect expressed preferences over trade policy. To the extent that public politics constrain trade policy, it would seem to be important to try to incorporate such considerations in our models.

Thinking about "Good" Trade Policy Virtually all of the work we have considered to this point has attempted to account for "bad" (i.e., welfare-reducing) trade policy. Given that the single most important fact of trade policy in the era since World War II is the historically striking trend toward greater trade liberalism and, by the mid-1980s, the truly extraordinary low level of overall protection (aside from agriculture) in the world's major trading nations, this focus seems a bit peculiar. If we believe, as surely we must, that political economy forces help explain the structure of protection, it seems equally clear that political economy forces must help explain this liberality. These forces need not be the same and, in fact, are unlikely to be the same. Thus the gains from new work on the domestic political economy of liberalization would seem to be sizable. One approach to this question might be to study the link between liberalization and protection. Whereas in the era of tariff politics, liberalization and protection were simply directions in a relatively unidimensional scale, in an era in which omnibus trade legislation endorses the power to liberalize at the same time it changes the terms of administered protection, it is not only difficult to say whether a piece of such legislation is protectionist or liberalizing, it seems likely that the two are related. Nevertheless, the extant models have brought us some distance in understanding the political economy of trade policy.

See also Heckscher-Ohlin model; nontariff measures; quotas; specific-factors model; tariffs

FURTHER READING

Anderson, Kym, and Robert Baldwin. 1987. "The Political Market for Protection in Industrial Countries." In *Protection, Cooperation, Integration, and Development,* edi

ted by Ali M. El Agraa. New York: Macmillan, 20 36. An excellent survey of empirical work on the political economy of trade policy.

Bauer, Raymond, Ithiel de Sola Pool, and Lewis Anthony Dexter. 1963. *American Business and Public Policy: The Politics of Foreign Trade*. Chicago: Aldine. A classic work on the political economy of trade policy that emphasizes, among many other things, the independence of legislators in the determination of trade policy using a framework that captures information transfer and communication.

Bhagwati, Jagdish, and T. N. Srinivasan. 1980. "Revenue Seeking: A Generalization of the Theory of Tariffs." *Journal of Political Economy* 88 (6):1069 87. Systematized and extended to the case of general equilibrium the argument that resource using political competition increases the costs of protection.

Blonigen, Bruce, and Thomas Prusa. 2003. "Antidumping." In *Handbook of International Trade*, edited by E. Kwan Choi and James Harrigan. Oxford: Blackwell, 251 84. A survey of contingent protection.

Bradford, Scott. 2006. "Protection and Unemployment." *Journal of International Economics* 69 (2): 257 71. In corporates both labor market search and unions into a political economy model with equilibrium unemployment.

Dixit, Avinash, Gene Grossman, and Elhanan Helpman. 1997. "Common Agency and Coordination: General Theory and Application to Tax Policy." *Journal of Political Economy* 105 (4):752 69. A particularly clear exposition of the Grossman Helpman model.

Findlay, Ronald, and Stanislaw Wellisz. 1982. "Endogenous Tariffs, the Political Economy of Trade Restrictions, and Welfare." In *Import Competition and Response*, edited by Jagdish Bhagwati. Chicago: University of Chicago Press, 223 34. Standard reference on the "tariff formation function" model of lobbying on the tariff. Make sure to read the comment by Leslie Young that follows the main paper.

Gawande, Kishore, and Pravin Krishna. 2003. "The Political Economy of Trade Policy: Empirical Approaches." In *Handbook of International Trade*, edited by Kwan Choi and James Harrigan. Oxford: Blackwell, 213 50. An excellent survey of empirical work on the political economy of trade policy.

Grossman, Gene, and Elhanan Helpman. 1994. "Protection for Sale." *American Economic Review* 94 (4): 833 50. By making both the state and groups active participants, this important paper embeds lobbying models that treat the state as passive and political response function models that make groups passive. This is probably the currently most widely used framework for theoretical and empirical research on trade.

Helpman, Elhanan. 1997. "Politics and Trade Policy." In *Advances in Economics and Econometrics: Theory and Applications*, vol. 1, edited by David Kreps and Kenneth Wallis. Cambridge: Cambridge University Press, 19 45. A survey of the political economy of trade policy with a theoretical focus.

Hillman, Arye L. 1982. "Declining Industries and Political Support Protectionist Motives." *American Economic Review* 72 (5): 1180 87. Standard reference on the political response function model of tariff formation.

Hiscox, Michael. 2001. *International Trade and Political Conflict: Commerce, Coalitions, and Mobility*. Princeton, NJ: Princeton University Press. A particularly thorough theoretical and empirical analysis of the link between factor mobility and the determination of equilibrium protection.

Krueger, Anne O. 1974. "The Political Economy of Rent Seeking Society." *American Economic Review* 64 (3): 291 303. Drawing on Tullock's analysis of rent seeking, this important paper develops an empirical analysis of the costs of protection including rent seeking costs.

Lowi, Theodore. 1964. "American Business, Public Policy, Case Studies, and Political Theory." *World Politics* 16 (4): 676 715. A classic work on the political economy of trade policy that might be the highest impact book review ever written. Lowi mediates the apparent conflict between Schattschneider and Bauer/Pool/Dexter by arguing that changes in the institutional definition of trade policy induced changes in the structure of politics.

Magee, Stephen. 1978. "Three Simple Tests of the Stolper Samuelson Theorem." In *Issues in International Economics*, edited by Peter Oppenheimer. Stocksfield, UK: Oriel Press, 138 53. This widely cited paper sought to test whether factors are intersectorally mobile (consistent with the Heckscher Ohlin model) or sectorally fixed (consistent with the Cairnes Haberler model and, more loosely, the Ricardo Viner model of fixed factors with a

single mobile factor) by examining the lobbying posi
tions of factor and labor on a trade bill. Magee found that
capital and labor tend to lobby together by sector.

Magee, Stephen, William Brock, and Leslie Young. 1989.
Black Hole Tariffs and Endogenous Policy Theory. Cam
bridge: Cambridge University Press. Magee, Brock, and
Young developed one of the first systematic lines of re
search on the political economy of trade policy. They
sought to integrate lobbying and electoral politics. This
book collects and extends this important early body of
research.

Matschke, Xenia, and Shane Sherlund. 2006. "Do Labor
Issues Matter in the Determination of U.S. Trade Pol
icy." *American Economic Review* 96 (1): 405 21.
Matschke and Sherlund argue theoretically and empir
ically that ignoring labor market structure in Grossman
Helpman empirics leads to faulty inference.

Mayer, Wolfgang. 1984. "Endogenous Tariff Formation."
American Economic Review 74 (5): 970 85. In addi
tion to a very clear explanation of the derivation of
preferences over policy from primitives of the economy,
this paper lays out the referendum model with great
clarity.

Mitra, Devashish. 1999. "Endogenous Lobby Formation
and Endogenous Protection: A Long Run Model of
Trade Policy Determination." *American Economic Re
view* 89 (5):1116 34. Develops a model in which the
decision of whether or not to join a group is en
dogenized. This analysis is then integrated into the
Grossman Helpman model.

Nelson, Douglas. 2006. "The Political Economy of Anti
dumping: A Survey." *European Journal of Political
Economy* 22 (3): 554 90. A survey of contingent pro
tection.

Olson, Mancur. 1965. *The Logic of Collective Action*. Bos
ton: Harvard University Press. Classic treatment of
group formation by rational agents when the goal of the
group is to provide a public good.

Peltzman, Sam. 1976. "Toward a More General Theory of
Regulation." *Journal of Law and Economics* 19 (2): 211
40. Original model of the political response function as a
basis for political economic analysis.

Rodrik, Dani. 1995. "Political Economy of Trade Policy."
In *Handbook of International Economics*, vol. 3, edited by
G. Grossman and K. Rogoff. Amsterdam: North

Holland, 1457 94. An often referenced survey of the
political economy of trade policy with a theoretical
focus.

Schattschneider, Edward Elmer. 1935. *Politics, Pressure, and
the Tariff*. Englewood Cliffs, NJ: Prentice Hall. A classic
work on the political economy of trade policy that em
phasizes the role of groups in lobbying for protection.

Tullock, Gordon. 1967. "The Welfare Costs of Tariffs,
Monopolies, and Theft." *Western Economic Journal* 5
(3): 224 32. Classic paper arguing that resources used in
the pursuit of a political goal must be counted in the cost
of that policy.

DOUGLAS NELSON

■ pollution haven hypothesis

A pollution haven is a country or region that attracts
pollution-intensive industry because its environ-
mental policy is less stringent than that of its trading
partners. A key issue in the debate over the effects of
trade on the environment is whether globalization
leads to the emergence of pollution havens: that is,
does trade and investment liberalization cause pol-
lution-intensive production to relocate from (mostly
high-income) countries with stringent environmen-
tal policy to (mostly low-income) countries with
weak or loosely enforced environmental regulations?
The pollution haven hypothesis posits that the an-
swer is yes.

If this hypothesis is correct, there is reason for
concern about the effects of trade on global envi-
ronmental quality. If production shifts from a region
with stringent environmental policy to one with
weak policy, the average pollution intensity of global
production rises, increasing global environmental
degradation. Moreover the shift in the incidence of
environmental harm from rich to poor countries
could increase human suffering, both because poorer
countries are less able to afford to respond to the
health effects of environmental damage and be-
cause poor countries are much more reliant on
natural capital (which is vulnerable to environmen-
tal damage) to sustain their income. Finally, an
exodus of pollution-intensive production from rich

to poor countries could lead to a political back-lash against stringent environmental policy. This possibility is sometimes referred to as a "race to the bottom" governments may weaken or fail to en-force environmental policies due to fear of job loss as firms threaten to leave for countries with weaker policy.

Discussions of the pollution haven hypothesis often confuse two different questions. One question is whether tightening environmental policy in a country (holding all else constant, including the trade regime) reduces net exports and/or net inward for-eign direct investment in pollution-intensive indus-tries. This question is simply whether environmental policy is one of the many factors that affect interna-tional competitiveness in pollution-intensive in-dustries. I will refer to this as the *competitiveness hypothesis*.

A second question is whether changing the trade regime by liberalizing trade or investment leads to a systematic reallocation of pollution-intensive pro-duction from countries with stringent environmen-tal policy to countries with weak environmental policy. For this to be true, something much stronger than the competitiveness hypothesis must be satis-fied: the effect of environmental policy on competi-tiveness has to be strong enough to be the key de-terminant of the overall pattern of trade and investment. I will refer to this as the *pollution haven hypothesis*.

Some authors distinguish between the two ques-tions by defining a pollution haven *effect* to exist if the competitiveness hypothesis is satisfied. The stronger pollution haven *hypothesis* then requires both that a pollution haven effect exists and that it be strong enough to determine the direction of trade and in-vestment flows.

Competitiveness and the Pollution Haven Ef-fect: Theory The prediction that tighter environ-mental policy will reduce competitiveness is quite robust and requires only one key assumption that more stringent environmental policy raises produc-tion costs. The most prominent alternative hypoth-esis is the Porter hypothesis. Porter and van der Linde (1995) argued that more stringent environmental

policy may actually increase productivity, for exam-ple by stimulating innovation. However most stan-dard economic models predict that placing increased constraints on firms (such as via tighter environ-mental policy) will raise costs and hence reduce competitiveness.

Competitiveness and the Pollution Haven Ef-fect: Empirical Evidence Testing the competitive-ness hypothesis has been very challenging (see Cope-land and Taylor 2004; and Copeland and Gulati 2006 for more detailed reviews). The most signifi-cant difficulty has been a lack of good data on the stringency of environmental policy. Pollution taxes are rarely used by governments, and other forms of environmental policy are difficult to quantify. A number of researchers have used county-level com-pliance with the Clean Air Act as a proxy for the stringency of environmental policy in the United States. The idea here is that states are required to have a plan to improve air quality in counties not in compliance with the act; consequently, noncompli-ance is taken to be a measure of relatively stringent environmental policy. Measures of abatement costs at the firm level have also been used; however, they are not available for very many countries, they are based on survey data, and they are problematic be-cause they are endogenous: if the pollution haven hypothesis is correct, the most polluting firms will have been driven away by tight pollution regulation, which would lead to lower measured average abate-ment costs in jurisdictions with stringent environ-mental policy (Keller and Levinson 2002).

A second set of difficulties arises from unobserved heterogeneity across jurisdictions. Trade and invest-ment are affected by many factors, and if some of these factors are correlated with the stringency of pollution policy, but not controlled for, then results will be affected by omitted variable bias. Recent work has used panel data techniques to deal with this type of issue; however the availability of such data is limited.

Endogeneity problems have also been a challenge. Pollution policy may respond directly to trade flows and investment and to variables correlated with trade and investment. As an example, suppose that

the competitiveness hypothesis is true but countries facing significant import pressure in pollution-intensive industries respond to trade liberalization by weakening environmental policy to help protect local firms from foreign competition. Then we could find that high imports are correlated with weak environmental policy, which is opposite to what the competitiveness hypothesis predicts. This type of problem has been dealt with by using techniques such as instrumental variables.

Early work testing the competitiveness hypothesis mainly used cross-sectional data to investigate how cross-industry differences in abatement costs affected trade or investment flows (for a survey, see Jaffe et al. 1995). Most of this work found that environmental policy had little or no effect. In some cases the estimated sign was opposite to what the competitiveness hypothesis predicted more stringent environmental policy was correlated with increased net exports. Some interpreted this finding as support for the Porter hypothesis; however, recent work has emphasized the role of omitted variable bias in affecting the results obtained from cross-sectional studies.

More recent work using panel data, mostly from the United States, has found support for the competitiveness hypothesis. A number of researchers have found that pollution regulations arising from the Clean Air Act had a significant effect on the location of new plants in pollution-intensive industries. Becker and Henderson (2000) found that new plant births in polluting industries in counties not in compliance with the Clean Air Act (and therefore subject to more stringent environmental policy) were 26 45 percent below levels in counties that were in compliance. There is also evidence that environmental policy affects trade flows. Levinson (1999) found that (all else being equal) high taxes on the disposal and processing of hazardous waste deter shipments of such waste from other states. Levinson and Taylor (2008) found that U.S. net imports are higher in industries with high abatement costs (once they accounted for the endogeneity of pollution policy using instrumental variables). This finding is consistent with the competitiveness hypothesis. The

results on foreign direct investment have been mixed. Keller and Levinson (2002) used a panel of U.S. state level data and found that states with relatively more stringent environmental policy (measured with an index of abatement costs per unit output) were less likely to attract foreign direct investment. So far, however, there is little evidence on whether the stringency of environmental policy affects the decisions of foreign firms to locate in developing countries.

Overall, a growing body of evidence supports the competitiveness hypothesis. Almost all of this work is based on U.S. data, however. More work, especially with data from developing countries, is still needed.

It is important to note that evidence that more stringent environmental policy reduces competitiveness in polluting industries does not imply that such policy is welfare decreasing. Weak environmental policy is an implicit subsidy for pollution-intensive production. Increased efficiency (and improvements in environmental quality) requires removal of this implicit subsidy, and this change will result in a shift from pollution-intensive production toward other cleaner production.

Pollution Haven Hypothesis: Theory The first pollution haven model was developed by Pethig (1976). He considered two countries that are completely identical except that one (North) has a higher pollution tax than the other (South). North's high pollution tax gives it a comparative advantage in the clean good. When trade is liberalized, trade shifts pollution-intensive production to the low-regulation country (South).

Copeland and Taylor (1994) showed how pollution havens could develop in poor countries even if all governments were free to choose whatever pollution policy best suited their countries. In the Copeland and Taylor model, countries are completely identical except in income levels. Northern countries are richer than Southern countries. The key assumption is that environmental quality is a normal good as income increases, consumers demand higher environmental quality. If government policy is responsive to consumer demand, then the model

predicts that richer countries will have more stringent environmental policy than poorer countries. Consequently, poor countries will have a comparative advantage in pollution-intensive production. When trade is liberalized, North exports clean goods, and South exports pollution-intensive goods: poor countries will become pollution havens.

This result is only part of the story. In reality, there are many reasons for trade, and the actual pattern of trade depends on the interaction among all such motives. Antweiler, Copeland, and Taylor (2001) developed a pollution haven model suitable for empirical testing by allowing for more than one motive for trade. In this model, countries differ in capital abundance as well as in income levels. Suppose that pollution-intensive industry is also capital intensive and that North is capital abundant. North's capital abundance tends to give it a comparative advantage in the pollution-intensive industry, but its stringent pollution policy tends to give it a comparative advantage in the clean industry. If the effects of North's capital abundance on the trade pattern are more important than the cost-increasing effect of its stricter environmental policy, then North will have a comparative advantage in the pollution-intensive good and trade liberalization will shift pollution-intensive production from the low-income South to the high-income North. In short, theory predicts that pollution policy is but one of many factors that affect trade. Whether or not trade liberalization leads to pollution havens depends on whether the effects of differences in environmental policy on production costs are more or less important than all of the other motives for trade.

Pollution Haven Hypothesis: Empirical Evidence Surprisingly, there is relatively little work that directly tests the pollution haven hypothesis. Several studies have found that the share of pollution-intensive goods in exports from developing countries has risen over time and the share of pollution-intensive goods in exports from members of the Organisation for Economic Co-operation and Development (OECD) has fallen over time. These trends are also consistent, however, with economic growth and capital accumulation in developing

countries and are not a full test of the pollution haven hypothesis.

Ederington, Levinson, and Minier (2004) looked at the pollution content of U.S. exports and imports during the period 1972 94 and found that U.S. imports from non OECD countries have become less pollution intensive over time relative to U.S. exports. This finding is opposite to what the pollution haven hypothesis predicts. Antweiler et al. (2001) estimate the effects of increased openness to trade on sulfur dioxide pollution, controlling for scale and other factors. They found that, all else equal, increased openness tended to increase SO_2 pollution in rich countries and reduce it in poor countries. Again this is opposite to what the pollution haven hypothesis predicts. Both Ederington et al. and Antweiler et al. argue that the evidence suggests that other factors (such as capital abundance or agglomeration) are more important than environmental policy in explaining trade flows.

In summary, there is a growing body of evidence that more stringent environmental policy does reduce a country's competitiveness in pollution-intensive industry, but there is little or no evidence to support the strong version of the pollution haven hypothesis: other factors are more important (on average) for affecting trade patterns. These are, however, the results of studies that look at broad patterns. There will be cases where poor countries have a comparative advantage in polluting industry for reasons unconnected with environmental policy (such as in polluting industries intensive in the use of unskilled labor). In cases such as this, weak pollution policy will reinforce the preexisting comparative advantage and increase the tendency for such industry to locate in a poor country.

Consumption-Generated Pollution Much analysis of the pollution haven hypothesis has focused on production-generated pollution. However, much pollution is generated by consumption (such as home heating, automobile transportation, and sewage). The effects of environmental policy on competitiveness are very different when regulations target consumption-generated pollution. For example, stringent automobile emission standards are applied

to all automobiles sold within a country, regardless of whether they are produced locally or imported. Consequently, tighter emission standards need not reduce domestic competitiveness. In fact, if it is easier for domestic producers than foreign producers to comply with stringent emission standards, then tighter environmental policy aimed at consumption-generated pollution may favor domestic producers over foreign producers. Hence the scope for trade-induced pollution haven effects is reduced when environmental policy affects consumption-generated pollution.

Natural Capital and Pollution Havens The effects of environmental policy on natural capital can be critically important for trade patterns. Consider the example of fisheries (where the fish stock is natural capital). Conservation requires that harvest rates be held below open access levels to ensure that the fish stock is sustainable. Hence in the short run, a country with weak conservation policy may harvest fish at a much greater rate than countries with good conservation policy. This can lead to a result analogous to pollution havens—trade liberalization would increase pressure on fish stocks in countries with weak conservation policy and exacerbate resource depletion in those countries.

However, it is possible for trade to have the opposite effect, especially in cases where resources are severely depleted. If a country has severely depleted its fish stock, then its potential supply of fish will be low. On the other hand, countries with relatively stringent conservation policies will have a much larger fish stock and hence a greater sustainable supply of fish. Hence when trade is liberalized, it is quite possible that countries with good conservation policies will be net exporters of fish and countries with poor conservation policies will be net importers of fish (Brander and Taylor 1997). Again, this is opposite to what the pollution haven hypothesis would predict.

See also location theory; trade and the environment

FURTHER READING

Antweiler, W., B. R. Copeland, and M. S. Taylor. 2001. "Is Free Trade Good for the Environment?" *American Economic Review* 91 (4): 877–90. An empirical study assessing the effect of trade on sulfur dioxide pollution.

Becker, R., and V. Henderson. 2000. "Effects of Air Quality Regulations on Polluting Industries." *Journal of Political Economy* 108 (2): 379–421. An influential study that investigates the effect of the Clean Air Act on industrial location.

Brander J. A., and M. S. Taylor. 1997. "International Trade between Consumer and Conservationist Countries." *Resource and Energy Economics* 19 (4): 267–98. A theoretical paper that shows how natural capital depletion affects trade flows.

Copeland, B. R., and S. Gulati. 2006. "Trade and the Environment in Developing Economies." In *Economic Development and Environmental Sustainability: New Policy Options*, edited by R. Lopez and M. A. Toman. Oxford University Press, 178–216. A nontechnical review of the literature on trade and the environment.

Copeland, B. R., and M.S. Taylor. 1994. "North South Trade and the Environment." *Quarterly Journal of Economics* 109 (3): 755–87. The first model of endogenous pollution havens.

———. 2004. "Trade, Growth, and the Environment." *Journal of Economic Literature* 42 (1): 7–71. A more technical review of the literature on both the competitiveness hypothesis and the pollution haven hypothesis.

Ederington, W. J., A. Levinson, and J. Minier. 2004. "Trade Liberalization and Pollution Havens." *Advances in Economic Analysis and Policy* 4 (2): Article 6. One of the few studies to explicitly test the pollution haven hypothesis.

Fullerton, D., ed. 2006. *The Economics of Pollution Havens.* Cheltenham, UK: Edward Elgar. An excellent collection of original research papers examining the pollution haven hypothesis.

Jaffe, A., S. Peterson, P. Portney, and R. Stavins. 1995. "Environmental Regulation and the Competitiveness of U.S. Manufacturing: What Does the Evidence Tell Us?" *The Journal of Economic Literature* 33 (1): 132–63. An influential review of the literature on the competitiveness hypothesis, but it came out before the recent work using panel data and confronting endogeneity problems.

Keller, W., and A. Levinson. 2002. "Pollution Abatement Costs and Foreign Direct Investment Inflows to U.S. States." *Review of Economics and Statistics* 84 (4): 691

703. Uses panel data to assess the effects of environ mental policy on inward foreign direct investment.

Levinson, A. 1999. "State Taxes and Interstate Hazardous Waste Shipments." *American Economic Review* 89 (3): 666 77. A test of the competitiveness hypothesis using a panel of U.S. data on hazardous waste flows.

Levinson A., and M. S. Taylor. 2008. "Trade and the En vironment: Unmasking the Pollution Haven Effect." *International Economic Review* 49 (1): 223 54. A test of the competitiveness hypothesis using a panel of U.S. trade data.

Pethig, R. 1976. "Pollution, Welfare, and Environmental Policy in the Theory of Comparative Advantage." *Journal of Environmental Economics and Management* 2 (3): 160 69. The first pollution haven model.

Porter, Michael E., and Claas van de Linde. 1995. "Toward a New Conception of the Environment Competitive ness Relationship." *Journal of Economic Perspectives* 9 (4): 97 118. Hypothesizes that environmental policy can increase competitiveness.

BRIAN R. COPELAND

■ poverty, global

The most commonly used definition of poverty is a lack of income; more specifically, having an income below an amount (the poverty line) deemed necessary for a minimum acceptable level of material well-being. Income is just a means to an end, however. There are more direct measures of the quality of life, such as health, education, and possibly freedom of expression. Poverty estimates also regularly make reference to mortality rates, schooling, and literacy.

Poverty may be defined and measured in either a relative sense (deprivation compared to societal norms) or an absolute sense (deprivation compared to being able to afford "objective" basic needs of food, clothing, and shelter). Discussions of global poverty have settled on an absolute definition of poverty, implicitly relegating concerns about income distribution to secondary importance. Doing so has deflected attention from redistribution and the role it can play in poverty reduction. Since the World Bank's 1990 *World Development Report*, "a dollar a day" has become the standard international poverty line a level based on an average of the poverty line being used in 10 developing countries at the time, although more recent estimates also use two dollars a day, which is a more relevant poverty line for middle-income countries such as those in Latin America.

Poverty estimates using the dollar-a-day line are made using "purchasing power parity," meaning they allow for differences in purchasing power as many items cost less in developing countries than in developed ones. The most commonly reported measure is the poverty headcount (the percentage of the population below the poverty line), but the same data can be used to calculate the poverty gap, which is a summary measure of the average distance of the poor below the poverty line, and the poverty severity index, which gives greater weight to the poorest. The poverty measures used in health and education are also absolute; that is, they do not base well-being on a comparison of the same indicators for the better off at either the national or international level.

Patterns of Poverty Estimates of dollar-a-day poverty are calculated only for the developing world. The proportion of absolutely poor in developed countries by this measure is nil or negligible. Figure 1 shows the evolution of dollar-a-day poverty since the early 1980s. The most striking trend is the dramatic fall in poverty in East Asia, powered largely by reductions in the number of poor in the world's most populous country, China, but assisted by more recent declines in neighboring Vietnam. There has been a slower, but still marked, decline in the poverty headcount in South Asia, including in the world's second largest country, India. By the mid-1990s the rate of decline was sufficient for the absolute number of poor people in South Asia to start falling. Asia thus makes up the vast bulk of poverty reduction in the closing decades of the last century. During the 1990s global poverty fell from 27.9 percent in 1990 to 21.1 percent in 2001, but excluding China these figures were 26.1 and 22.2 percent, respectively. That is, more than half the decrease in poverty came from China alone. Indeed sub-Saharan Africa, which has suffered economic hardship since the 1970s, saw a rise in income poverty, with close to half the people

Poverty head count: percentage of population below the poverty line

Absolute number in poverty (millions)

Legend:
East Asia and Pacific
East Europe and Central Asia
Latin America and the Caribbean
Middle East and North Africa
South Asia
Sub-Saharan Africa
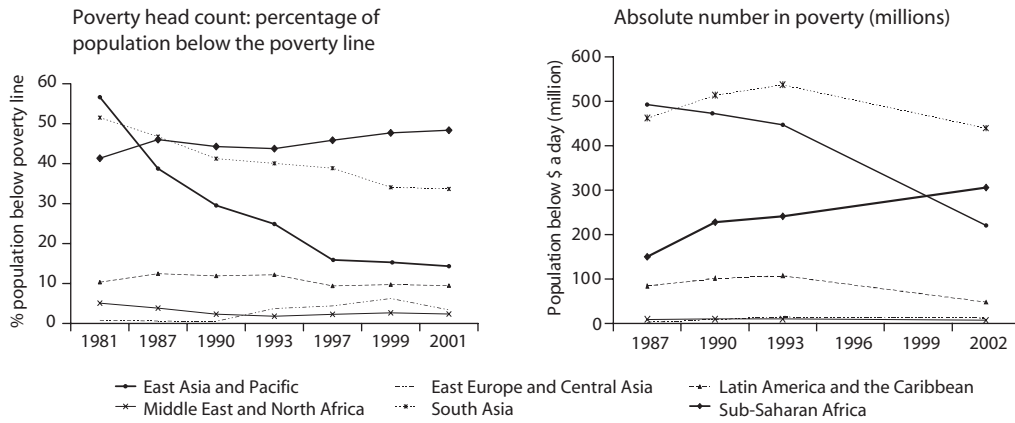
Figure 1

Dollar a day poverty by region, 1983 2002. Source: World Bank Poverty Monitoring Web site, http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTPAME/0,,contentMDK: 20188187~pagePK:210058~piPK:210062~theSitePK:384263,00.html.

on the subcontinent now living on less than a dollar a day. Africa's poor performance lies behind the low overall decrease in the number of poor from 1.20 billion in 1987 to 1.03 billion in 2002. Africa's share in this total rose from 12 percent to 30 percent over this period (whereas Asia's fell from 80 percent to 64 percent). The final trend of note is the resurgence of poverty in the formerly centrally planned economies following the collapse of the communist governments, though there has been some remission since the beginning of the 21st century.

The Africanization of poverty is also evident when considering other poverty measures. Health is commonly measured by infant mortality (the number of children who die before their first birthday per 1,000 live births) and child mortality (deaths between first and fifth birthdays per 1,000 children); these two indicators are combined to make under-five mortality. The United Nations Population Division reports population data, including mortality, from 1950, with projections to 2050, though only those to 2010 are shown here. The positive news is that mortality rates have been falling across the world (figure 2), although some African countries experienced a reversal in the 1990s as a result of HIV/AIDS and worsening health systems after three decades of economic decline. At the extreme, life expectancy in

Botswana has fallen by 29 years since 1990, from 64 to just 34 years. In Africa as a whole the decline in mortality rates has been insufficient to keep up with population growth, so the number of deaths has continued to increase. A turnaround is expected, but another expectation is that the situation in Africa will improve less than that elsewhere, so the continent will account for close to two-thirds of the world's under-five deaths in the coming decades.

Progress in education has gone in stops and starts. During the colonial period there was little formal education for most of the population in much of the developing world other than that provided by mission schools. Newly independent governments undertook substantial expansions of the education sector at all levels, though enrollments remained well below half of all children, especially in rural areas. Since 2002 there has been a renewed emphasis on raising primary education levels with the international Fast Track Initiative. Many countries, notably in Africa, have pushed for higher enrollments, scrapping school fees at the primary level, resulting in rapid increases in student numbers, although across the subcontinent as whole about 30 percent of children do not attend school, and many of those who do learn little. In Asia also large strides have been made in moving toward full enrollment in recent years so
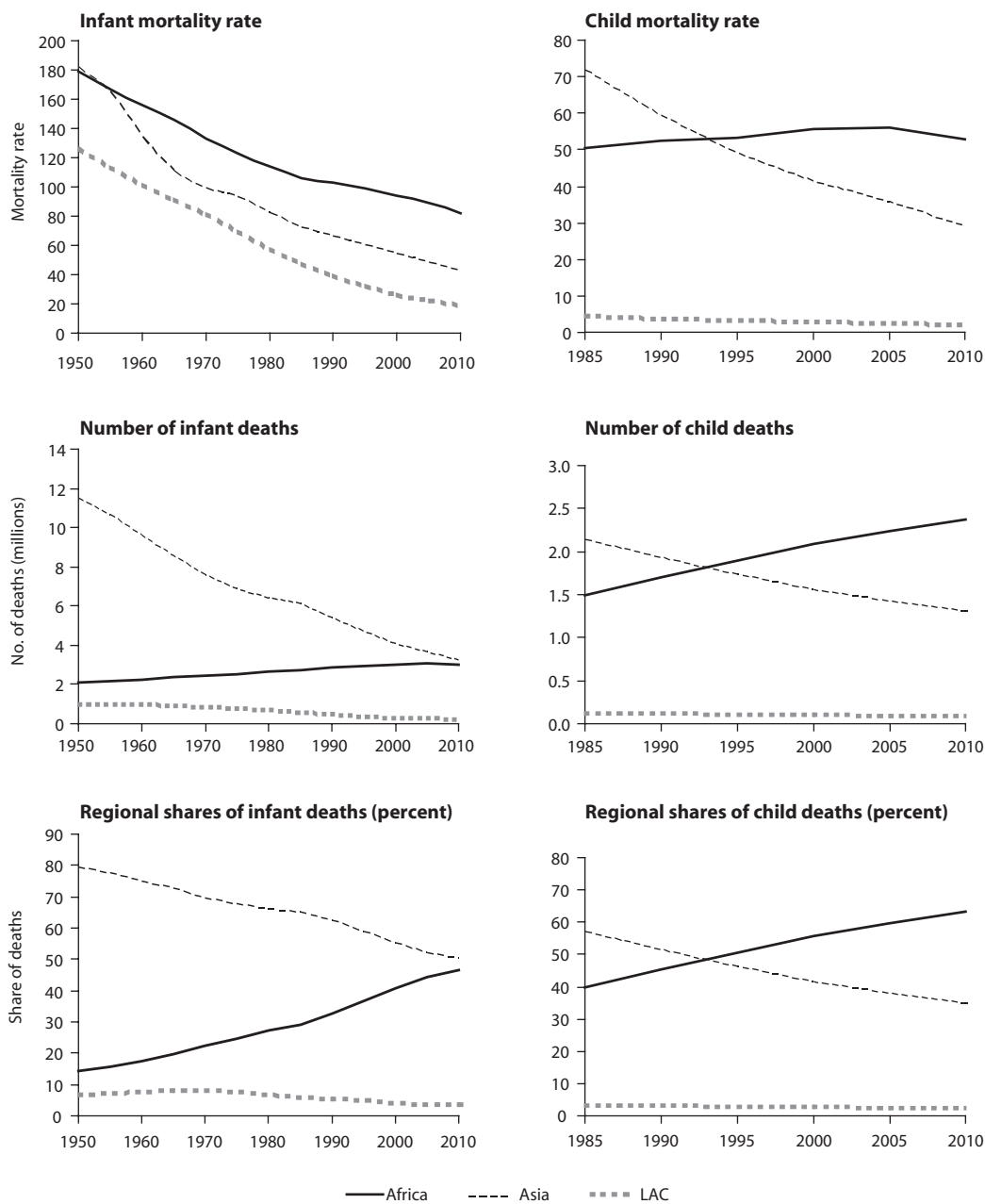
Figure 2

Trends in infant and child mortality. Source: UN Population Division database, http://esa.un.org/unpp/.

that primary enrollments exceeded 90 percent as of 2007, and the focus has shifted to secondary education with international agencies financing stipend schemes for girls in rural areas.

These aggregates hide systematic patterns in poverty at the national level. Poverty is nearly always higher in rural areas than urban ones: at least 80 percent of Africa's poor live in rural areas, where income opportunities are fewer and service provision lower. There are usually disadvantaged regions where poverty is higher: for example, infant and child mortality rates may be twice the national average. In Bangladesh, for instance, neonatal mortality (death in the first month) is 32 per 1,000 live births in Barisal Division, but 63 per 1,000 in the more conservative Sylhet Division, where tradition prevents women from seeking medical assistance even in the case of birth complications. These variations may reflect ethnic or racial biases and are not confined to developing countries: African Americans have a higher rate of under-five mortality than do white South Africans (indeed, the United States as a whole has a higher under-five mortality rate than does Cuba).

Gender also plays an important role in poverty. Girls are less likely than boys to be in school, and women continue to earn less than men for their labor. In rural India a woman's daily agricultural wage is typically half that of a man's. Female-headed households are typically poorer than those that are male-headed, although this is not the case when there is an absent male sending remittances back to the household. And families in most of Asia prefer sons to the extent that female infants are often neglected, with food and health care given preferentially to male children.

Globalization of Poverty The globalization of poverty is the view that we are today affected by the poverty of those in other nations in ways we were not in the past. In the words of the former German premier Willy Brandt, who led the team responsible for the 1980 report *North-South*, "Ignoring the problems of the South is like a person on a sinking ship saying 'I'm all right, my end's not sinking.'" Brandt's view was one of global Keynesianism: growth in the developing world will power growth at home through increased demand for our exports, whereas decline overseas will cause eventual decline at home. Other aspects of global poverty are also thought to spill over to affect the West in ways that did not happen before, such as disease and the increased flow of refugees fleeing from conflict or disasters.

These views overstate the extent to which global interdependence is a new phenomenon. One-quarter of the Irish population emigrated to the United States during the potato famine of the 1840s, and a similar percentage of the Scottish population did so as a result of Highland clearances in the same century (landlords turning land cultivated by smallholding tenants over to sheep grazing). Earlier settlers brought with them illnesses that wiped out a large part of the indigenous population. Centuries earlier the Black Death spread from Asia to Europe, killing up to two-thirds of the European population.

Strategies to Tackle Global Poverty Two broad approaches may be identified in the dominant discourses on development. Official agencies emphasize the role that growth can play and believe that growth is best brought about through liberalization of markets. Others, including most nongovernmental organizations, place greater emphasis on the need for development aid.

According to the first view, the best support the West can give to reduce poverty is to encourage developing countries to deepen their engagement with the global economy and to open up to private sector capital flows. But critics point to the apparent hypocrisy of developed countries, which continue to practice protective policies, notably subsidies to agriculture, that are harmful to the welfare of developing countries. For example, U.S. subsidies to domestic cotton producers reduce the prices received by farmers growing cotton in countries such as Brazil and Zambia. Furthermore, although there can be benefits from capital inflows, short-term speculative flows can be destabilizing, as shown by the East Asian crisis of the late 1990s. Others point out that an increased global flow of labor, rather than capital, would probably do most to reduce poverty but is restricted by the developed countries. On balance,

however, one can say that many developing countries engaged in excessive degrees of state intervention following independence and that dismantling these controls has freed up resources to be used more productively in growth. But this argument does not support the view that full liberalization of all markets is the most conducive means to sustained growth and certainly not the fastest route to poverty reduction.

Aid's impact on poverty has been muted since poverty reduction has not always been the explicit focus of foreign aid. This is partly because donor countries also use aid to pursue their own political and commercial objectives. But it is also because aid has focused on other issues, such as the development of infrastructure in the 1950s and 1960s and on restoring economic growth, albeit with mixed success, in the 1980s. There was a clearer poverty focus in the 1970s, under the banner of Basic Needs, and it reemerged more strongly in the 1990s. Reviews of aid's poverty impact provide a mixed picture there are undoubted success stories of aid improving the lives of poor people, but a lot of aid at best leaves them unaffected and at worst harms them. Efforts to ensure that aid indeed benefits the poor are being made but can run into political opposition in both developed and developing countries.

The latest stage in efforts to focus aid on poverty are the Millennium Development Goals (MDGS), formerly the International Development Targets. These are a set of targets based on various UN resolutions, such as the halving of income poverty by 2015, a two-thirds reduction in infant and child mortality by the same date, and the already missed target of equality in girls' education by 2005. Aid donors have responded to this increased poverty focus by concentrating aid on poorer countries and reorientating it toward basic services, such as providing money for rural health clinics rather than urban hospitals and the stipend schemes for girls to attend secondary school in Asia mentioned earlier.

In most low-income countries aid is given in the context of Poverty Reduction Strategy Papers (PRSP), which are a formal requirement for countries receiving debt relief from the Highly Indebted Poor Country Initiative. Central to these strategies, promoted by the World Bank and the International Monetary Fund, is the notion that economic growth is the key to poverty reduction and that growth is best brought about by liberalized markets: hence critics argue that PRSPs are just "business as usual." A contrary view of development comes from the United Nations Development Program (UNDP) through its *Human Development Reports*, which argues that there is a larger role of government through direct investments in social infrastructure, which can bring about reductions in poverty faster than is possible by growth alone.

Future Poverty Trends Analysis of likely future trends shows a continuation of the patterns discussed here, most notably the continuing Africanization of poverty as poverty levels fall in Asia but with at best slow progress in Africa. The MDG for a halving of income poverty will be met in most of Asia, but not in Africa or in the former communist countries of Eastern Europe; poverty in these regions began to rise after the MDG base year of 1990 and is unlikely to meet its former low levels by 2015, let alone fall to half the 1990 levels.

These forecasts assume "business as usual." But it is likely that things will be worse than that on account of both manmade and natural disasters. Poor countries and poor people are least able to withstand these shocks; households do not have insurance and the country is less able to afford reconstruction as shown comparing the amounts spent on reconstruction after Hurricane Katrina in 2005 and after the far worse Indian Ocean tsunami in 2004. Conflict destroys livelihoods, disrupts education, allows the reemergence of disease, and can set growth back by years. Civil wars and regional conflicts continue to have these adverse effects around the world. A larger shock from a global conflict is not an impossibility, which would render all projections meaningless.

See also development; evolution of development thinking; global income inequality; International Monetary Fund (IMF); World Bank

FURTHER READING

Black, Richard, and Howard White, eds. 2004. *Targeting Development: Critical Perspectives on the Millennium*

Development Goals. London: Routledge. Critiques each of the MDGs and what governments in both the developed and developing world can do to achieve them.

UNDP. Annual. *Human Development Report.* New York: UNDP. Focuses on a different development related issue each year and updates various development and poverty measures, such as the Human Poverty Index.

World Bank. Annual. *Global Monitoring Report.* Washington, DC: World Bank. Covers MDG related indicators, with summary overview and a different thematic focus each year.

HOWARD WHITE

■ preferential trade area

See free trade area

■ pricing to market

See exchange rate pass-through

■ primary products trade

Primary products are unprocessed raw materials. They include agricultural, forestry, fishing, and mining products, including minerals and fuels. Unlike the processing of these raw materials into manufactured goods, the ability to produce primary products is limited by natural endowments: some regions have the deposits of oil or minerals, the soil and climate, or the natural fisheries that make it possible to obtain these products, while other regions do not. As a consequence, human beings have engaged in primary product trade from time immemorial. Today, primary products account for around 25 percent of global merchandise trade. The ability to exchange agricultural and fisheries products has enriched diets, while access to forestry and mining products has enhanced the consumption possibilities of communities located far from the sources of these products.

Despite the advantages of primary product trade to consumers, it has been the subject of intense criticism. Do the countries producing these products gain equivalent benefits? Does concentrating a country's resources in the production of primary products enhance the welfare of its citizens? Does it provide a sound basis for a country's long-term growth and development? These questions are of more than academic concern for developing countries, many of whose exports are heavily concentrated in primary products. Indeed, today primary products account for only 19 percent of the merchandise exports of high-income countries, but 36 percent of the exports of middle-income countries and 50 percent of the exports of low-income countries. For many developing regions, this figure is even higher: 69 percent in sub-Saharan Africa and fully 80 percent in the Middle East and North Africa.

Classical and Neoclassical Trade Theory Classical economists were the first to systematically address whether primary exports were good for welfare, which they answered in the affirmative. Based on differences in techniques of production, the Ricardian trade model showed that there were potential gains for each country engaging in comparative advantage trade. In such trade, each country specializes in those products it can produce at lower relative labor cost than other countries. The Heckscher-Ohlin trade model of the 1920s assumed countries could choose from the same array of technologies to produce any product, but they still faced differences in their factor endowments and therefore in relative factor prices. These differences resulted in different choices of technology and in different relative product costs. If each country chose to specialize in the products in which it had a lower opportunity cost, global production would be carried out by the most efficient producers, and global output would increase. Trade would allow consumers in all countries to increase their consumption beyond the limits imposed by their own capabilities.

Prebisch-Singer Critique The Heckscher-Ohlin model and its variants, generally referred to as the neoclassical trade model, showed that every country gains from comparative advantage trade, no matter what specific product a country exports. This conclusion was based, however, on the assumption that

markets are perfectly competitive and that relative prices therefore reflect relative costs. The first serious critique of the neoclassical model came in the 1930s and 1940s, from the economists Raúl Prebisch and Hans Singer, who challenged this fundamental assumption. They argued that the benefits for primary exporters were limited. Primary exporters faced both fluctuating prices for their products and declining terms of trade, resulting in volatile and diminishing gains over time.

The most commonly cited elements of their critique have to do with the elasticities involved. First, price inelastic demand for agricultural products combines with frequent weather- or pest-induced shifts in supply to create considerable volatility in agricultural prices. Second, income inelastic demand for many primary products, especially agricultural products (Engel's Law), combined with considerably higher income elasticities for manufactures, contributes to declining relative prices or commodity terms of trade for many primary exporting countries. In addition, the development of synthetic substitutes can contribute to sharp declines in primary product prices. Synthetic dyes, chemical fertilizers, and synthetic fibers, for example, have caused collapses in the markets for natural dyes, for organic and mineral fertilizers, and for natural fibers, while plastics have affected the markets for a broad range of forestry and mineral products. Critics of primary export specialization point to numerous episodes of economic collapse in primary-exporting countries resulting from these developments.

In addition, Prebisch and Singer first advanced the argument that markets were not perfectly competitive and that international market prices might in fact not reflect relative costs of production. Although the production of many primary products could be said to occur under competitive conditions, producers faced a limited number of buyers (oligopsony or monopsony) when they sold their raw materials. These buyers had the market power to drive producer prices below their competitive level. On the other hand, manufactured goods purchased by developing countries tended to be produced in oligopolistic markets, where suppliers had the market power to

potentially raise prices above competitive levels. Put together, these market characteristics suggested that primary-exporting developing countries got paid too little for their exports, while paying too much for their imports. A corollary is that technological progress would be encouraged in the industrial countries but discouraged in the primary exporting developing countries. More recent trade models have included elements of noncompetitive markets, especially increasing returns to scale, which capture these aspects of international pricing that disadvantage primary exporters (e.g., Greenaway 1991).

The experience of the Latin American region in the century following independence in the 1820s provided Prebisch with clear evidence of the volatility of primary product export earnings, as countries experienced related boom-and-bust cycles with alarming frequency. More recently, commodity crises have affected producers of coffee, cocoa, tin, and cotton, to name but a few. Prebisch also found evidence for his arguments in the data collected by the United Nations, which showed a steady decline in the prices of primary products compared with manufactures in international markets from 1876 to 1938. More recent studies have found similar declines over the 20th century for non-oil-exporting developing countries.

This original empirical evidence was critiqued, however, because it did not take account of changes in the costs of production. Productivity gains in primary production could have accounted for the relative decline in primary product prices. Given the difficulty of calculating the "double-factoral" terms of trade that take account of productivity changes, the debate over the empirical evidence continues today.

Agricultural Subsidies and Developed-Country Trade Barriers What is not subject to debate is the effect that agricultural subsidies have had on global prices for many products. For decades, the advanced industrial countries have provided large subsidies to their agricultural producers, especially producers of grains, dairy, cotton, and tobacco. These subsidies have given developed-country agricultural products a price advantage that has lowered global prices of these products and undercut less subsidized producers in developing countries (Oxfam 2002). The

decline in international prices for these products has also been encouraged by persistent protectionist policies in the industrial countries, as well as by efforts to encourage increased production of agricultural exports in developing countries under structural adjustment programs. Developing country protests regarding these pricing issues led to the breakdown of the World Trade Organization's Doha Round of negotiations.

Distribution of Gains from Trade Another element of the critique of primary product exports is the question of the distribution of any gains within the producing country. The Stolper-Samuelson extension of the neoclassical model predicts that trade will raise the price of the factor utilized intensively in the exported product. Many observers assumed that for developing countries this factor would be labor, since capital was scarce, and that as a result international trade would benefit the poor through increased wages. Primary products exports increase the demand for natural resources much more than for labor, however, so most of the benefits of such trade accrue to landowners, often an already wealthy elite. It is no surprise, then, to find that the century of primary export specialization and trade in Latin America, from the 1820s to the 1920s, created a region with the most unequal income distribution in the world, a legacy that persists to this day. The increasing concentration of natural resource ownership over this period ensured that most of the benefits from primary export trade went to a small land-owning elite. In more recent times, many developing countries have found that much of the wealth from exploitation of natural resources has accrued to foreign investors.

If there are overall gains from trade, it is possible to distribute the gains broadly, although whether this occurs in practice depends on the social and political institutions that influence access to natural resources. One approach is to maintain public ownership of key natural resources and to use the revenues for public expenditures that provide widespread benefits to the population. Unfortunately, this is not always the way these revenues are actually used. Another approach is to allow private development of natural resources but tax resource owners in order to spread the benefits

of trade. These efforts have met with mixed success due to the ability of foreign investors to influence political outcomes. In Guatemala, for example, the United Fruit Company, a large landowner and the country's main banana exporter, was able to avoid all but minimal taxes for more than 50 years.

Dynamic Trade Benefits Beyond these issues regarding the static gains from trade, perhaps of even greater significance for the developing countries is the effect of primary product specialization and trade on the long-term growth of the economy: the "dynamic" trade effect. Proponents of comparative advantage trade argue that it can serve as an "engine of growth" for the overall economy. This will occur through expanded factor endowments and spillovers or linkages.

Expanded factor endowments will occur if the export sector is able to attract foreign investment and the capital, skilled labor, and technological know-how that it can bring into the country. Human capital may expand as a result of the "learning-by-doing" effect. These effects will increase productivity and incomes in the export industries, creating profits that can be reinvested into continued productivity gains and expansion of the sector.

At the same time, spillovers and linkages may develop to other productive activities in the economy. The income from primary exports increases consumer demand and may finance investment in new industries. Backward linkages may develop to domestic suppliers of inputs for the export industry, as demand for those inputs grows and allows suppliers to produce at efficient levels. Forward linkages may develop to the processing, transportation, storage, and marketing of the export products, with spillovers to services for other domestic products. Infrastructure constructed to facilitate the production and trade of the export products may also serve other productive sectors. Exports can therefore serve as a growth pole, stimulating other economic activities in an ever-expanding process of economic diversification.

Limited Dynamic Benefits for Primary Product Exporters While advocates of comparative advantage trade predict that dynamic benefits will accrue to

any country, regardless of the specific products they export, critics argue that the extent of these benefits can be limited for countries whose comparative advantage lies in primary products. Indeed, some have gone so far as to posit a "resource curse" effect, whereby abundant natural resources reduce a country's growth rate.

There are numerous theories as to why this may be the case. Factor endowments will not expand if declining terms of trade reduce profitability of primary exports. The same negative impact on investment can occur if extraction of nonrenewable resources raises production costs or if renewable resources are overtaxed, so that the country's comparative advantage gradually declines.

Even if investment in the primary sector continues to be robust, critics argue that spillovers and linkages are unlikely to develop, for many reasons. Some arguments focus on the employment and income effects of resource wealth. Rural wealth and employment may raise industrial real wages and therefore reduce incentives to invest in manufacturing. At the same time, primary products may generate fewer learning effects than manufactures, so that the country's labor force skills remain underdeveloped.

On the other hand, some primary products may be land- and capital-intensive, generating limited employment and therefore limited consumption multiplier effects. This is most likely with mining activities, but can also occur with large-scale, mechanized agricultural production. While these activities can create considerable wealth for investors and landowners, this wealth may not remain in the developing country, especially in the case of multinational investment. Profits, dividends, royalties, and the earnings of skilled expatriate employees may be remitted rather than entering into domestic financial institutions.

One reason for this capital flight may be the continued low profitability of manufacturing investment in the country, as spillovers fail to materialize. Economies of scale in manufacturing, especially in the case of technologically complex production processes, can make it difficult for developing countries to produce on a profitable scale. This can limit

backward linkages to capital and intermediate goods, forward linkages to processing, or consumption linkages to consumer durables, especially for smaller countries. These cost disadvantages can be compounded by the limited technological capabilities of the labor force and the underdeveloped state of the country's institutions and infrastructure. While there may be spillovers from infrastructure developed to support primary production, these extractive activities are often geographically isolated from main population centers, their infrastructure therefore conferring few external benefits.

Geographic isolation is not the only reason why centers of primary production may become "enclaves." The very success of primary exports can, paradoxically, have this effect, in a process that has been dubbed the "Dutch disease." A booming export sector can act as a black hole, pulling resources into itself and thereby raising production costs for domestic manufacturers, while the inflow of foreign currency causes the prices of competing manufactured imports to fall. Rather than creating positive spillovers, the booming export sector creates negative externalities for other tradables sectors. Resource windfalls have also been found to have a negative effect on overall savings and investment and to stimulate rent-seeking behavior.

There is evidence to support the contention that primary exporters do not gain dynamic trade benefits. Sachs and Warner (2001), for example, found that countries with a higher percentage of natural resource based exports in 1970 grew more slowly from 1970 to 1990. There are, nevertheless, exceptions to this general rule, as proponents of the "staples theory" argue (Lewis 1989). Malaysia and Thailand, for example, are resource-rich countries that have experienced exceptionally high growth rates since the 1970s (Reinhardt 2000). Recent literature suggests that the impact of resource abundance on growth is not immutable but rather depends on institutions, educational and industrial policy, and the nature of the learning process (Mehlum, Moene, and Torvik 2006). On the other hand, there is evidence that these variables are themselves influenced by natural resource abundance, being indirect channels through

which resource wealth inhibits economic growth (Papyrakis and Gerlagh 2004).

While the debate over primary exports continues, trade liberalization in the past several decades has increased the importance of this issue for many developing countries, who find their short-term welfare and long-term prospects tied to the fortunes of their primary products. Although commodity price booms such as those of recent years have created windfall gains for many developing countries, they still face the difficult task of turning those primary export gains into a sound basis for growth and development.

See also North-South trade; terms of trade; trade and economic development, international

FURTHER READING

Atkinson, Giles, and Kirk Hamilton. 2003. "Savings, Growth and the Resource Curse Hypothesis." *World Development* 31 (11): 1793 1807. Shows link between resource wealth and low rates of savings and growth.

Auty, Richard. 1994. "Industrial Policy Reform in Six Large Newly Industrializing Countries: The Resource Curse Thesis." *World Development* 22 (1): 11 26. Case studies that demonstrate the "resource curse" argument.

Dodaro, Santo. 1991. "Comparative Advantage, Trade, and Growth: Export Led Growth Revisited." *World Development* 19 (9): 1153 65. Demonstrates importance of industrial exports to growth.

Greenaway, David. 1991. "New Trade Theories and Developing Countries." In *Current Issues in Development Economics*, edited by V. N. Balasubramanyam and Sanjaya Lall. New York: St. Martin's Press, 156 70. Reviews literature incorporating noncompetitive models.

Lewis, S. R. 1989. "Primary Exporting Countries." In *Handbook of Development Economics*, vol. 2, edited by Hollis Chenery and T. N. Srinivasan. Amsterdam: North Holland Press, 1541 99. Detailed review of primary trade theory and country examples showing how developing countries can benefit from it.

Mehlum, Halvor, Karl Moene, and Ragnar Torvik. 2006. "Institutions and the Resource Curse." *Economic Journal* 116 (508): 1 20. Suggests that the resource curse is a matter of policy rather than necessity.

Oxfam. 2002. *Rigged Rules and Double Standards: Trade, Globalization, and the Fight against Poverty.* London: Oxfam International. Estimates costs to developing countries of developed country agricultural subsidies and tariffs on primary imports.

Papyrakis, Elissaios, and Reyer Gerlagh. 2004. "The Resource Curse Hypothesis and Its Transmission Channels," *Journal of Comparative Economics* 24 (1): 181 93. Empirical study demonstrating the mechanisms by which resource wealth can slow growth.

———. 2006. "Resource Windfalls, Investment, and Long term Income." *Resources Policy* 31 (2): 117 28. Shows how primary export windfalls can lower investment and long run growth.

Prebisch, Raúl. 1950. *The Economic Development of Latin America and Its Principal Problems.* Lake Success, NY: United Nations. Prebisch's classic paper outlining his theory of primary product trade.

Reinhardt, Nola. 2000. "Back to Basics in Malaysia and Thailand: The Role of Resource Based Exports in Their Export Led Growth." *World Development* 28 (1): 57 77. Shows how some countries have benefited from primary product trade.

Sachs, Jeffrey, and Andrew Warner. 1995. "Natural Resource Abundance and Economic Growth," NBER Working Paper No. 5398. Cambridge, MA: National Bureau of Economic Research. Reviews the arguments and evidence against primary product trade.

———. 2001. "The Curse of Natural Resources." *European Economic Review* 45 (4 6): 827 38. Empirical study showing the negative relationship between primary product exports and developing country growth.

Singer, Hans. 1950. "The Distribution of Trade between Investing and Borrowing Countries." *American Economic Review* 40 (2): 473 85. Singer's classic paper critiquing developing country primary product trade.

NOLA REINHARDT

■ private sector involvement

See bail-ins

■ protectionism

See nontariff measures; tariffs

■ proximity-concentration hypothesis

The proximity-concentration hypothesis or trade-off is a way to explain why firms undertake horizontal foreign direct investment (FDI), in which firms produce the same product at plants in multiple countries. Firms do so when the benefits of avoiding high trade costs outweigh the costs incurred when producing in multiple markets. Firms are more likely to undertake horizontal FDI as transport costs and trade barriers become higher and as investment barriers and scale economies at the plant level become lower.

Horizontal FDI models are motivated by a large and increasing share of FDI, especially two-way FDI, which takes place between economies with similar factor proportions. Models of vertical FDI (in which different parts of the production process are done in different countries), by contrast, suggest that FDI should arise between economies with different factor proportions (and in only one direction). The proximity-concentration hypothesis also draws from the literature on ownership, location, and internalization advantages as drivers of FDI (mostly ownership and location advantages).

Understanding why multinationals often establish multiple plants in various countries is important because it highlights factors that affect these decisions and informs businesses and policymakers on appropriate policy measures that could influence such investments.

Theoretical Framework A firm has several modes of servicing an overseas market: exports, FDI, or licensing. In deciding on its overseas expansion strategy, a firm faces trade-offs between proximity and concentration advantages. Proximity advantages could arise from savings on transportation, tariff, and warehousing costs, superior customer services, better understanding of the local market conditions, and access to specialized inputs. The concentration advantage arises from any scale economies at the firm or plant level. Economies of scale at the firm level are associated with activities such as research and development, marketing, advertising, and management services, while economies of scale at the plant level result in lower per unit cost as the scale of production increases. Brainard (1997) states, "The proximity-concentration hypothesis predicts that firms should expand horizontally across borders whenever the advantages of access to the destination market outweigh the advantages from production scale economies."

Among the many models of horizontal FDI, Markusen (1984) pioneered one in which multiplant economies are a driver for multinational investment. Multiplant economies are the result of a joint input that can be shared across multiple production facilities and whose productivity is not affected by the number of facilities at which it is used. In this scenario, multinational production avoids costly duplication of the joint input that would arise with independent national firms.

Brainard (1993) coined the term *proximity-concentration trade-off hypothesis*. She uses a model in which firms chose between exports and FDI based on scale economies at the firm and plant level, and transportation costs that increase with distance. Absent factor price differences, firms compare the additional variable cost of exporting with the additional fixed cost of setting up a plant overseas. This model has three possible equilibria: a multinational equilibrium where all firms are multinationals with plants in both countries; a trade equilibrium where all firms are national with a single plant and export to foreign countries; and a mixed equilibrium where multinational and national firms coexist.

With a two-stage production structure, two-way horizontal intraindustry FDI completely crowds out trade in goods even without differences in factor proportions, as long as proximity advantages are strong relative to concentration advantages. For intermediate ranges of transportation costs and scale economies at the firm level relative to plant-level scale economies, there is a mixed equilibrium in which national and multinational firms coexist in this sector: there is two-way trade in final goods and two-way FDI. For a fixed number of firms, the proportion of firms that export is greater the larger the plant-level fixed costs and the smaller the transportation and other trade costs and the size of each market.

If concentration advantages dominate in upstream activities (activities that take place at the initial stages of production, e.g., production of intermediate inputs) and proximity advantages dominate downstream activities (activities that take place after the final stage of production through the consumption stage, e.g., transportation, distribution, and marketing activities), the incorporation of a third stage of production, such as sales, results in two-way intrafirm, intraindustry trade in intermediate goods replacing trade in final goods; here multinational activity is complementary to trade in goods. When both factor proportions differences and proximity-concentration trade-off are combined, firms decide whether to export or undertake overseas production based on the relative strength of these considerations.

In the knowledge capital model (see Markusen 1997; Markusen and Maskus 2001; Carr et al. 2001) both horizontal and vertical multinationals can arise depending on country characteristics such as size, size differences, relative differences in endowments, trade costs, and investment costs. Horizontal multinational firms are associated with similarities between countries in both size and relative factor endowments. The model suggests that affiliate production and trade tend to be substitutes for similar countries and complements for countries with substantially different relative factor endowments.

An extension of the horizontal model introduces intraindustry firm heterogeneity in the proximity-concentration framework to examine the trade-off between exporting and FDI (Helpman et al. 2003). Exporting involves lower sunk costs but higher per-unit costs relative to FDI. Every industry has heterogeneous firms that differ in productivity. Firms with higher productivity are more profitable in all three activities: production for the domestic market, exports, and FDI.

The least productive firms are not profitable and exit the market. Low-productivity firms serve only the domestic market as they expect to incur losses from exports or FDI. We then have a range of higher productivity firms that can serve both domestic and foreign markets. The less productive among these

expect to be profitable through exporting, but anticipate losses from FDI. Hence they undertake exports but not FDI. The most productive firms, on the other hand, chose FDI over exports as they expect greater operating profits from FDI than from exporting. Thus, while proximity-concentration variables determine the level of firm productivity needed to make international expansion attractive, the intraindustry distribution of firm productivity plays an important role in explaining the composition of trade and the firm's choice between exports and FDI.

Empirical Validation In empirical tests of the proximity-concentration hypothesis, we expect to find the share of affiliate sales in total foreign sales to be positively related to distance/trade costs, local market size, and firm-level economies of scale, and negatively related to plant-level economies of scale. Several papers provide some empirical support for this hypothesis. Using industry-level data from U.S. multinationals, Brainard (1997) finds that the proximity-concentration hypothesis is fairly robust in explaining the share of total sales accounted for by exports as compared with affiliate sales. The share of affiliate sales increases as income similarities between trading partners, transportation costs, foreign trade barriers, and firm-level scale economies increase, and decreases as foreign investment barriers, foreign taxes, and scale economies at the plant level increase.

Ekholm's (1998) analysis, using Swedish firm-level multinational data, finds some support for the proximity-concentration hypothesis: concentration advantage negatively affects the likelihood of affiliate production, while an increase in transportation costs/ geographic distance increases the share of affiliate sales in total foreign sales. Carr et al. (2001) find results consistent with Brainard (1997): outward investment increases with the sum of the economic sizes of the source and host countries, their similarity in size, the relative skilled-labor abundance of the parent nation, and the interaction between size and relative endowment differences. Using firm-level U.S. multinational data, Helpman et al. (2004) find

firms substitute FDI for exports where transportation/tariff costs are high and plant-level returns to scale are relatively weak. When firm-level heterogeneity is introduced, however, greater firm-level heterogeneity in productivity leads to significantly more subsidiary sales relative to exports and the size of this effect is the same order of magnitude as that of trade frictions/costs.

Policy Implications The proximity-concentration hypothesis provides explanations for FDI between countries with very similar factor proportions, cross-border multiplant configurations, intraindustry trade, and complementarities between trade and investment. Several policy implications arise from proximity-concentration models. Local governments can undertake policy initiatives such as tariffs, subsidies, or taxes to influence the type of multinational activity undertaken in their country. Brainard (1993) suggests that the relative structure of tariffs on intermediate and final goods influences firms' decision regarding what to produce at home and what to produce overseas. Where there are relatively large proximity advantages and scale economies at the firm level, restrictions on investments may have more damaging welfare effects than restrictions on trade.

The proximity-concentration hypothesis suggests some factors that influence whether a firm decides to establish plants in both countries or to serve the foreign market by exports. There is broad empirical support for the theory that firms do indeed weigh the benefits of proximity against the costs of forgoing the economies of scale that would be enjoyed by concentrating production.

See also foreign direct investment: the OLI framework

FURTHER READING

Brainard, S. Lael. 1993. "A Simple Theory of Multinational Corporations and Trade with a Trade off between Proximity and Concentration." NBER Working Paper No. 4269 (February). Cambridge, MA: National Bureau of Economic Research. The first paper to introduce the term *proximity concentration trade off.*

———. 1997. "An Empirical Assessment of the Proximity Concentration Trade off between Multinational Sales and Trade." *American Economic Review* 87 (4): 520 44. An empirical examination of the proximity concentration hypothesis.

Carr, David, James Markusen, and Keith Maskus. 2001. "Estimating the Knowledge Capital Model of the Multinational Enterprise." *American Economic Review* 91 (3): 693 708. Empirical test of the predictions of the knowledge capital model on the relationship between foreign affiliate sales and country characteristics.

Ekholm, Karolina. 1998. "Proximity Advantage, Scale Economies, and the Location of Production." In *The Geography of the Multinational Firm*, edited by Pontus Braunerhjelm and Karolina Ekholm. Amsterdam: Kluwer Academic Press, 59 76. An empirical paper that examines among other things the proximity concentration trade off.

Helpman, Elhanan, Marc J. Melitz, and Stephen R. Yeaple. 2003. "Export versus FDI." NBER Working Paper No. 9439 (January). Cambridge, MA: National Bureau of Economic Research.

———. 2004. "Export versus FDI with Heterogeneous Firms." *American Economic Review* 94 (1): 300 16. Both papers by Helpman et al. introduce intraindustry firm heterogeneity in the proximity concentration framework and examines the trade off between exporting and FDI.

Markusen, James R. 1984. "Multinationals, Multiplant Economies, and Gains from Trade." *Journal of International Economics* 16 (3 4): 205 26. One of the early papers explaining horizontal FDI in the presence of multiplant economies.

———. 1997. "Trade versus Investment Liberalization." NBER Working Paper No. 6231 (October). Cambridge, MA: National Bureau of Economic Research. Uses the knowledge capital framework to examine the relationship between trade and investment liberalization.

Markusen, James R., and Keith E. Maskus. 2001. "General Equilibrium Approaches to the Multinational Firm: A Review of Theory and Evidence." NBER Working Paper No. 8334 (June). Cambridge, MA: National Bureau of Economic Research. A useful review of the literature on multinational firms.

USHA NAIR REICHERT

■ purchasing power parity

The notion of purchasing power parity (PPP) has a long intellectual history and can be traced to the 16th-century writings of scholars from the University of Salamanca in Spain. The modern definition of PPP, usually credited to Gustav Cassel (1918), is quite intuitive: when measured in the same unit, the monies of different countries should have the same purchasing power and command the same basket of goods. Otherwise, international arbitrage should bring about adjustments in prices, exchange rates, or both, which will ultimately restore parity. Another way to interpret the parity condition is that the exchange rate between two currencies should equal the ratio of the countries' price levels.

Despite its simplicity, the parity condition is the subject of many empirical studies, driven mainly by its significant implications for the global economy. For instance, PPP is a major building block of most models in international economics. The relevance of these models and their policy implications thus depends critically on the validity of PPP. Another application of PPP is the comparison of national income levels. Some economists believe that a meaningful comparison of income across countries should be based on, instead of market exchange rates, PPP exchange rates that control for price differentials of same goods across countries.

The use of PPP goes beyond academic interest, however. In the context of the global economy, exchange rate misalignment is a main source of imbalances in trade and capital accounts. These imbalances, if left unchecked, can create intense stresses for both individual economies and the global system. The PPP condition describes the relationship between exchange rates and national price levels and is commonly used as a benchmark for evaluating exchange rate misalignment.

The absolute and the relative PPP conditions are the two commonly discussed versions of PPP. The absolute PPP is given by

$$s = p - p*$$

where s is the exchange rate expressed as the domestic price of the foreign currency, p is the domestic price index, and p* is the corresponding foreign price index. All the variables are in logs. The relative version is given by

$$s = c + p - p*$$

where c is constant.

Under the absolute PPP, movements in the relative price, p p*, and in the exchange rate offset each other to maintain the parity. On the other hand, the relative PPP is a less stringent condition and requires only that the proportion of exchange rate variations is the same as the proportion of variations in the relative price.

Law of One Price and PPP The law of one price (LOP) states that prices of identical goods from different locations are the same after adjusting for exchange rates. International arbitrage is the main argument behind the LOP. The difference between LOP and PPP is that the former concept refers to the actual price level of a good while the latter refers to an index of prices for different goods. LOP is a main building block of PPP. For instance, it can be shown that the PPP condition can be derived when only one good satisfies LOP (Niehans 1984). The requirement of only one good, instead of all the goods, satisfying LOP is an interesting result and it implies a high likelihood that PPP holds.

Most studies that evaluate the LOP condition find that, in general, the condition does not hold and deviations from the LOP vary substantially across product types. These studies, however, usually employ price subindexes, such as the price of chemical products. Thus they are examining the relative rather than the absolute LOP. Studies that use actual individual prices are far less common. The documented price convergence in the European car market during the integration process is one of the few pieces of evidence that are in favor of LOP (Goldberg and Verboven 2005). Overall, while one is hard pressed to find favorable evidence among the empirical studies, a casual look at, say, the global precious metal market suggests that the presence of one good satisfying the LOP may not be an extremely stringent condition.

Basics of Evaluating PPP Arguably, the PPP is the most intensely examined parity condition in international economics. Constrained by data limita-

tions, empirical studies usually examine the validity of relative instead of absolute PPP. Specifically, price indexes do not allow an easy comparison of absolute price levels across countries.

The choice of price indexes is an issue in evaluating PPP. The common choice is between consumer price and producer (wholesale) price indexes. Some researchers use the former index as a proxy for the price index of nontradable goods and the latter one as a proxy for tradable goods. In general, the use of producer price index yields stronger evidence in favor of PPP than the use of consumer price index does. Other candidates include price indexes of imports and exports, the GDP deflator, and components of the GDP deflator.

The sample period used in PPP studies ranges from the post-1973 floating period to historical samples that cover 100 years or more. There is a trade-off between short and long samples. An advantage of long historical samples is that they are better suited for studying long-term trends and thus for evaluating the long-run PPP. A drawback, however, is that these samples cover periods of different exchange regime arrangements and significant economic changes that may complicate the analysis. Compared with long historical samples, the post-1973 period may not have enough observations to reveal reversion to the PPP. Nonetheless, results from the post-1973 period are derived from a more homogenous setting and are likely to be more relevant for current policy considerations.

Real Exchange Rate Persistence The persistence of real exchange rates is commonly used to infer the validity of PPP. If PPP holds continuously, the real exchange rate is constant. It is not difficult to verify that the instantaneous PPP does not hold since a constant real exchange rate is typically not observed in reality. This is why most empirical exercises focus on long-run PPP. Under the long-run PPP regime, deviations from parity are possible but short-lived. Over time, the relative price and the exchange rate adjust to restore the parity condition and the corresponding real exchange converges to its equilibrium value. Thus an operational interpretation of long-run PPP is that real exchange rates are mean-reverting,

and the test for the validity of PPP can be translated to a test for mean-reverting behavior of real exchange rates.

The mean-reverting behavior of real exchange rates and, hence the PPP condition, is commonly assessed using unit root test procedures. Indeed, the evolution of these empirical studies closely tracks the development of unit root test procedures. In the 1980s, the introduction of the Dickey-Fuller test revolutionized the assessment of economic data persistence in general and real exchange rate persistence in particular. Since then, a flurry of studies have used various versions of unit root tests, including the original augment Dickey-Fuller (ADF) tests, improved versions of ADF tests, Bayesian unit root tests, the fractional integration test, panel data unit root tests, and procedures allowing for alternative adjustment mechanisms to evaluate real exchange rate persistence.

Traditionally, real exchange rate persistence is evaluated in a linear time-series framework. If the real exchange rate follows a nonlinear path, then the use of a linear model can overstate its degree of persistence. A few nonlinear models, including models with structural breaks, fractional integration models, Markov switching models, and threshold autoregressive models, have been applied to real exchange rate data. In general, allowance for nonlinear dynamics enhances the ability to reveal the reversion to the PPP and lowers the empirical estimate of real exchange rate persistence (Michael, Nobay, and Peel 1997).

On balance, the empirical evidence suggests that PPP tends to hold in the long run but not in the short run. Complementing academic empirical findings, foreign exchange dealers, who jointly determine exchange rates in the global market, also indicate that PPP provides a good gauge of exchange rate movements only in the long run (Cheung and Chinn 2001).

Of particular importance is the survivorship bias in PPP analysis (Froot and Rogoff 1995). This bias refers to the common practice of investigating long historical data from developed countries. The practice mainly reflects data availability rather than

research interest. One implication of the bias is that results from developed countries may overstate the empirical support for PPP. A related issue is whether developed and developing countries have similar real exchange rate behavior. Cheung and Lai (2000a) conduct a large-scale analysis and find substantial cross-country heterogeneity in the persistence of deviations from parity, and it is more likely, rather than less likely, to find parity reversion for developing countries than developed countries.

A perplexing and well-known empirical regularity is that real exchange rates display both a high level of persistence and an intense amount of short-term volatility (Rogoff 1996). If one attributes large short-term volatility to some dominating nominal shocks, then the observed persistence is too high to be explained by price stickiness. This phenomenon is labeled a "PPP puzzle." Several attempts have been made to explain the puzzle. One study shows that, when sampling uncertainty is taken into consideration, the estimated persistence parameter gives only a very imprecise measure of the true persistence (Cheung and Lai 2000b). Thus the observed persistence puzzle may not be a well-defined puzzle. Subsequent studies also point out that the estimated level of persistence can be reduced to a "nonpuzzling" level when, say, nonlinearity adjustment mechanisms are considered.

Economic Factors So far the discussion has mainly drawn from studies based on the time-series (à la intertemporal) properties of exchange rate and price data. What economic factors affect the behavior of the real exchange rate, which is a measure of PPP deviations?

The Balassa-Samuelson effect is perhaps the most discussed economic force shaping real exchange rate behavior. The hypothesis is related to the observation that, when measured in the same unit, price levels in high-income countries are higher than those in low-income countries. The price differential is due to the difference in productivities in the tradables and nontradables sectors. The Balassa-Samuelson effect is a supply-side factor. A demand-side factor is government spending. It is perceived that nontradables account for a large fraction of government spending. Thus government spending tends to have a positive impact on a country's real exchange rate.

Other economic factors that affect real exchange rates include net foreign asset position, distribution sector, and market structure. Since the turn of the 21st century, the net foreign asset position has been introduced to a real exchange rate equation via portfolio-balance channels (Lane and Milesi-Ferretti 2002). The role of net foreign asset position is related to the effect of current/trade account balances articulated in the 1980s and 1990s. The effect of net foreign asset position is verified in a number of empirical studies. It is noted, however, that the construction of data on national net foreign assets involves approximations and these data are all in U.S. dollars.

The prices faced by consumers are affected by distribution costs. The productivity and efficiency of the distribution sector will have an impact on the price structure and, thus, on real exchange rate behavior. Intuitively, the distribution activity occurs within a locale and is likely to be a "nontradable" service. In the Balassa-Samuelson effect parlance, improved distribution sector productivity should be associated with a depreciation of the real exchange rate. Nonetheless, the extant empirical evidence points in a different direction   a result that may be driven by tradable components in the distribution sector (MacDonald and Ricci 2005). One caveat is the paucity of data on productivities of distribution sectors.

The implication of market structure for the adjustment of relative prices to exchange rate movements was recognized in the 1980s. A model with, say, a markup pricing strategy that is a function of the degree of monopolistic power offers an easy illustration of the market structure effect on real exchange rate behavior. Indeed, the market structure has implications for both the persistence and volatility of real exchange rates. As always, the availability of proper data is a challenge for the empirical valuation of the market structure effect. Since the turn of the 21st century, some positive evidence of market

structure effects has been reported using sectoral data (Cheung, Chinn, and Fujii 2001).

Inter- and Intracountry Analysis Usually, both LOP and PPP are stated with the implicit assumption that prices in different countries are being compared. There is no reason, however, that the two parity conditions should not apply to prices of the same goods within a country. Indeed, the study of relative price behavior within a country avoids a few problems that plague the cross-country analysis.

For instance, the procedures for compiling price data and constructing indexes can differ across countries. Within a country, however, price data are likely to be collected and recorded in one unified system. Thus data compatibility should not be a significant issue in comparing the prices of same goods across cities and regions within a country.

There are a few other reasons why results based on intracountry data are easier to interpret than those from intercountry data. First, intracountry relative prices are subject to similar fiscal and monetary policies, whereas intercountry relative prices are subject to possibly different policies. Second, the trade barriers within a country are usually less severe than those across countries. Third, the intracountry relative prices are literally linked via an exchange rate that is credibly fixed at the one-to-one level and, thus, are free from uncertain exchange rate variability. Indeed, Engel and Rogers (1996) have shown that there is an enormous increase in relative price variability across national borders. These factors imply that intracountry relative prices of identical goods are more likely to meet the parity condition.

Given these considerations, it is perhaps not too surprising to find that intracountry data, compared with intercountry data, display a faster convergence rate and are more likely to satisfy the parity condition. The result is supported by studies using either individual goods prices or price indexes.

See also Balassa-Samuelson effect; equilibrium exchange rate; exchange rate forecasting; exchange rate volatility; interest parity conditions; nontraded goods; real exchange rate

FURTHER READING

Cassel, Gustav. 1918. "Abnormal Deviations in International Exchanges." *Economic Journal* 28: 413 15. Brings the PPP to its current context.

Cheung, Yin Wong, and Menzie Chinn. 2001. "Currency Traders and Exchange Rate Dynamics: A Survey of the U.S. Market." *Journal of International Money and Finance* 20: 439 71. One of a series of articles using survey responses to examine issues in exchange rate economics including PPP.

Cheung, Yin Wong, Menzie Chinn, and Eiji Fujii. 2001. "Market Structure and the Persistence of Sectoral Real Exchange Rates." *International Journal of Finance and Economics* 6: 95 114. Explicitly documents the effects of market structure on real exchange rate behavior.

Cheung, Yin Wong, and Kon S. Lai. 2000a. "On Cross Country Differences in the Persistence of Real Exchange Rates." *Journal of International Economics* 50: 375 97. Reports different country groups display different mean reverting real exchange rate dynamics and different persistence patterns.

———. 2000b. "On the Purchasing Power Parity Puzzle." *Journal of International Economics* 52: 321 30. Shows that the PPP puzzle may not be well founded because data only give a very imprecise estimate of real exchange rate persistence.

Cheung, Yin Wong, Kon S. Lai, and Michael Bergman. 2004. "Dissecting the PPP Puzzle: The Unconventional Roles of Nominal Exchange Rate and Price Adjustments." *Journal of International Economics* 64: 135 50. Reports the perplexing result that, compared with relative prices, exchange rates display a slower speed of adjustment toward equilibrium.

Chinn, Menzie D. 2000. "The Usual Suspects? Productivity and Demand Shocks and Asia Pacific Real Exchange Rates." *Review of International Economics* 8: 20 43. Documents the presence of the Balassa Samuelson effect in a group of Asia Pacific economies.

Engel, Charles, and John H. Rogers. 1996. "How Wide Is the Border?" *American Economic Review* 86: 1112 25. The first paper that quantifies the effect of a national border on real exchange rate volatility.

Froot, Kenneth A., and Kenneth Rogoff. 1995. "Perspectives on PPP and Long Run Real Exchange Rates." In *Handbook of International Economics*, edited by Kenneth

Rogoff and Gene Grossman. Amsterdam: North Hol land, 1647 88. An often cited survey article with a lot of references.

Goldberg, Pinelopi K., and Frank Verboven. 2005. "Mar ket Integration and Convergence to the Law of One Price: Evidence from the European Car Market." *Jour nal of International Economics* 65: 49 73. Presents evi dence in favor of LOP.

Lane, Philip, and Gian Maria Milesi Ferretti. 2002. "Ex ternal Wealth, the Trade Balance, and the Real Exchange Rate." *European Economic Review* 46: 1049 71. Con nects real exchange rate behavior to net foreign asset positions.

MacDonald, Ronald, and Luca Antonio Ricci. 2005. "The Real Exchange Rate and the Balassa Samuelson Effect: The Role of the Distribution Sector." *Pacific Economic Review* 10: 29 48. Develops a theoretical model and provides empirical evidence of the link between the distribution sector and real exchange rate dynamics.

Michael, Panos, A. Robert Nobay, and David A. Peel. 1997. "Transaction Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation." *Journal of Political Economy* 105: 862 79. Appealing to the effect of transaction costs, the study uses nonlinear models to describe real exchange rate behavior.

Niehans, Jurg. 1984. *International Monetary Economics*. Baltimore, MD: Johns Hopkins University Press. Pre sents a model that gives the PPP condition with the assumption of one product satisfies the LOP condition.

Officer, Lawrence H. 1982. *Purchasing Power Parity and Exchange Rates: Theory, Evidence, and Relevance*. Greenwich, CT: JIA Press. An extensive review of the early literature on PPP.

Rogoff, Kenneth. 1996. "The Purchasing Power Parity Puzzle." *Journal of Economic Literature* 34: 647 68. An often cited survey article with a lot of references.

Sarno, Lucio, and Mark P. Taylor. 2002. "Purchasing Power Parity and the Real Exchange Rate." *IMF Staff Papers* 49: 65 105. Covers some recent developments in the empirical PPP literature.

YIN WONG CHEUNG

■ quantity theory of money

The classical equation of exchange as it is applied today equates the money supply multiplied by the velocity of circulation (the number of times the average currency unit changes hands in a given year) with the economy's gross domestic product (the quantity of output produced multiplied by the price level). Although the American economist Irving Fisher originally included transactions of previously produced goods and assets as part of the output measure, only newly produced goods are included in present-day calculations of gross domestic product. If the velocity of circulation and output are both constant, there is a one-to-one relationship between a change in the money supply and a change in prices. This yields a "quantity theory of money" under which a doubling of the money supply must be accompanied by a doubling of the price level (or, when the rate of money supply growth doubles, the rate of inflation also doubles). The Nobel Prize winning economist Milton Friedman (1956) restated this quantity theory of money by modifying the assumption of constant velocity, allowing velocity to adjust in response to changes in expected inflation and the returns available on other assets such as bonds and equities.

The tendency for velocity to rise as expected inflation rises reflects people's incentive to unload a depreciating currency before its purchasing power erodes. This reinforces the effects of faster money supply growth on inflation because, at the same time that more money is being printed, people want to hold even less of it than they did before the inflationary process began. There is, in fact, an exact inverse relationship between the proportion of income held as money and the velocity of circulation with, say, a halving of average money holdings requiring that the currency circulate at double the old rate in order to buy the same goods as before. In the extreme case of "hyperinflation" (broadly defined as inflation exceeding 50 percent per month), money holdings plunge and the inflation rate significantly outstrips the rate of money supply growth. Conversely, under deflation, velocity is likely to fall: people hold onto their money longer as they recognize that the same funds will buy more goods and services over time if prices continue to decline. Thus money demand rises as money supply falls, again exaggerating the effects of the money supply change on the aggregate price level.

The demand for money is also influenced by the economy's output level. If output rises, money demand should rise too, as increased production generates more income and spending power. This would put downward pressure on prices. Just as sustained inflation is possible only when the rate of money growth rises above the rate of growth of output (and money demand), so too does sustained deflation occur under conditions of insufficient money growth. According to the monetarist school, which emphasizes the importance of the money supply as a long-run determinant of prices and gross domestic product, stable prices could be achieved by simply tying the rate of money supply growth to the

long-run rate of growth of output. This would keep money demand and money supply in balance so long as velocity is stable.

Critics, such as those of the Keynesian school, which views the effects of monetary policy as unreliable and uncertain in practice, argue that such a policy rule would have undesirable consequences because velocity historically has been subject to fluctuations that require offsetting movements in the money supply. On the other hand, Friedman (1960) argued that velocity has been unstable only because policy has been unstable and that adoption of a constant growth rate rule for the money supply would keep inflation expectations, along with velocity, steady.

Friedman's restatement of the quantity theory of money tempers the classical prediction that money supply increases exert one-to-one effects on prices even in the short run. Under the restatement, velocity is allowed to adjust to a new level if the inflation environment changes. However, once real money holdings have adjusted to their new lower equilibrium level following a rise in inflation, the rise in velocity should end with a one-to-one relationship between money growth and inflation. Such a pattern is typically observed over extended periods. Moreover Friedman's (1992) famous proposition that inflation is always and everywhere a monetary phenomenon receives support from the fact that it is hard, if not impossible, to find any episodes of significant inflation that have not been accompanied by accelerating rates of monetary expansion. The importance of this link is further illustrated by the fact that, while a one-third money supply reduction in the eastern Confederate States in April 1864 dramatically reduced inflation there, inflation continued to run rampant in the western portion where the monetary cutback was postponed.

Criticisms of Quantity Theory Critics of the quantity theory approach have questioned not only the stability of velocity and money demand but also the determination of the money supply itself. That is, a link between money and prices does not necessarily prove that money supply movements are driving price movements. The economists Sargent and Wallace (1973) argue that such "reverse" causation could arise when a government is dependent on the revenue earned by inflating away the value of its outstanding money issues. Here, higher inflation produces faster rates of monetary expansion, and not the other way around. As higher expected inflation leads individuals to reduce their real money balances, this reduces individuals' exposure to the inflation tax and reduces the government's revenue from inflation. To keep inflation tax revenue at its old level, the government must accelerate the rate of monetary expansion so as to increase the inflation tax rate and offset the decline in the inflation tax base as real money balances fall. This novel, albeit controversial, perspective could not apply in the more usual situation where governments are able to finance their expenditures through conventional taxes and bond issues.

The quantity theory's predictions have also been questioned by adherents of "backing theory." Under this view, it is the quality rather than the quantity of money that matters. Whereas unbacked paper money may well be as inflationary as the standard quantity theory assumes, this need not be true of money that is credibly backed by future taxes or other provisions for their future retirement from circulation. Pioneering analysis of the American colonies prior to the Revolutionary War by Smith (1985) suggests that even large money issues might be willingly held rather than spent   implying a fall in the velocity of circulation and an absence of the inflationary pressures predicted by the quantity theory provided that they were properly backed. The case of Maryland stands out because that colony undertook to accumulate funds set aside for future purchases of pounds sterling that would retire the colony's paper money at a predetermined rate of exchange. Data limitations make it hard to conclusively determine the practical extent to which backing reduced the inflationary effects of the Maryland currency issues and those of other colonies, but recent work suggests that Pennsylvania enjoyed the long-run constancy of velocity, and proportional relationship between money and prices, implied by the quantity theory (Grubb 2005).

Excess Monetary Creation Whatever predictive power the quantity theory approach may have over the long term, even its strongest adherents would accept that short-run dynamics and adjustments militate against a one-for-one relation between money and prices in the short run. Excess money creation will eventually lead to inflation, but extra liquidity may well initially push down interest rates and encourage greater output and employment. Until these beneficial effects are reversed, the extra money issue may "buy" at least the illusion of greater prosperity. This, in turn, may produce a temptation to inflate. While the continued ratcheting up of the money supply will eventually lead to hyperinflation, such an extreme outcome usually occurs only when a government finds itself unable to obtain funding from any source other than the printing press. Such episodes, while unfortunate, have nevertheless provided economists with ample opportunity to observe not only the inflationary consequences of such rampant excess money growth but also the surge in the velocity of circulation as individuals become progressively less willing to hold the depreciating currency.

See also debt deflation; Federal Reserve Board; money supply; purchasing power parity; seigniorage; time inconsistency problem

FURTHER READING

Burdekin, Richard C. K., and Marc D. Weidenmier. 2001. "Inflation Is Always and Everywhere a Monetary Phenomenon: Richmond vs. Houston in 1864." *American Economic Review* 91 (5): 1621 30. Illustrates how divergent inflation performance in the eastern and western Confederacy was linked to a reform measure that reduced the money supply in the eastern Confederacy while initially leaving the money supply unchanged in the west.

Fisher, Irving. 1922. *The Purchasing Power of Money: Its Determination and Relation to Credit, Interest, and Crises.* New York: Macmillan. Fisher's original exposition of the equation of exchange and quantity theory of money.

Friedman, Milton, ed. 1956. *Studies in the Quantity Theory of Money.* Chicago: University of Chicago Press. Includes Friedman's own restatement of the quantity theory and Phillip Cagan's pioneering analysis of money demand under hyperinflation.

———. 1960. *A Program for Monetary Stability.* New York: Fordham University Press. Lays out Friedman's case for a constant growth rate rule for monetary policy.

———. 1992. *Money Mischief: Episodes in Monetary Policy.* New York: Harcourt Brace Jovanovich. Highly accessible general analysis of the causes and consequences of monetary expansion, both past and present, can be found in chapters 2 and 8.

Grubb, Farley. 2005. "Two Theories of Money Reconciled: The Colonial Puzzle Revisited with New Evidence." NBER Working Paper No.11784. Washington, DC: National Bureau of Economic Research. Available at http://www.nber.org/papers/w11784. Revisits the applicability of the quantity theory of money to the American colonial experience and finds Pennsylvania data to be consistent with the quantity theory's predictions.

Sargent, Thomas J., and Neil Wallace. 1973. "Rational Expectations and the Dynamics of Hyperinflation." *International Economic Review* 14 (2): 328 50. Points to the potential dependence of money supply movements on price movements when a government must support its operations through the inflation tax.

Smith, Bruce D. 1985. "Some Colonial Evidence on Two Theories of Money: Maryland and the Carolinas." *Journal of Political Economy* 93 (6): 1178 1211. Suggests that even large money supply increases need not be inflationary if the money is backed by future tax revenue or other provisions for the retirement of the currency.

RICHARD C. K. BURDEKIN

■ quotas

Quotas in international trade are government-imposed maximum amounts of a product that can be imported into or exported from a country during a specific period of time (usually a year). The maximum can be a quantitative limit specifying the maximum tons or units or square yards that can be traded, or it can be a maximum in value terms, denominated in either domestic or foreign currency. A *global* import quota sets a single maximum amount

that can be imported, regardless of the country of origin. Import quotas are frequently subdivided among exporting countries, specifying a maximum amount that can be imported from each trading partner.

Use of Quotas Governments have used import and export quotas to achieve a number of different policy objectives. Governments often used import quotas to protect domestic producers from import competition or to encourage domestic production. An example would be the import quotas on petroleum in the United States from 1959 to 1973, intended to promote domestic exploration for oil and oil production to enhance national security. The United States relied on import quotas to prevent lower-priced imports from undermining price supports for domestic agricultural products, or from increasing the cost of the support program. Countries with fixed exchange rates and escalating balance of payments deficits often resorted to import quotas to limit payments for imports and prevent large losses in international reserves. In a balance of payments "crisis," governments often chose quotas because of the certainty that a quota ceiling would limit imports.

Governments sometimes impose export quotas to ensure sufficient supplies of a product for domestic consumption or processing. Romania, for example, used export quotas on logs, lumber, and plywood to ensure a supply of inputs for domestic furniture producers. Another objective is to keep domestic prices of sensitive goods below world market prices. The United States, for example, imposed export quotas on logs in the mid-1970s in an attempt to prevent increases in the price of lumber from driving up the cost of homebuilding. For countries that are major suppliers of an export commodity, export quotas reduce supply to the world market and keep world market prices high. An example would be the Organization of the Petroleum Exporting Countries' (OPEC's) export quotas on petroleum. Voluntary export restraints (VERs) are export quotas that were imposed by an exporting country on shipments of a good to a particular importing country *at the request of the importing country.* They allowed importing countries to protect struggling industries without

violating international trading agreements by imposing their own import restrictions.

Administration of Import and Export Quotas Governments have used many different rules or mechanisms to determine which international traders effectively receive the right to import or export under quota systems.

"First come, first served." One method of administering a quota restriction is to simply allow traders to import or export the restricted product starting at the beginning of the quota period until the quota ceiling is reached and then close the border to additional trade for the rest of the time period. This method can cause a chaotic situation as traders "stampede to the border" at the beginning of the quota period. Countries that attempted to administer quotas under a first-come, first-served rule in the 1930s found that quota limits were often exceeded because information flows to the authorities were too slow to close the border quickly enough to prevent the quota from being overshipped. Governments began to allocate the rights to import under quota to firms or individuals according to sets of rules or allocation mechanisms.

Allocation by "traditional trade shares." This method allocates shares of the quota among traders on the basis of their trade volumes during a specified base period. A firm that had imported 10 percent of total imports during the historical base period would be allocated 10 percent of the total amount allowed in under the quota restriction. This method excludes traders who did not happen to be trading in the base period, so governments often reserve a small portion of the quota for "new" or "nontraditional" importers or exporters. The United States Department of Agriculture allocated import quotas for cheese to importing firms on the basis of historical trade shares.

Allocation in proportion to domestic production. During the period in which the United States used import quotas to restrict petroleum imports, import licenses were allocated to domestic refineries in proportion to refinery output (Bohi and Russell 1978). McCulloch and Johnson (1973) discuss the efficiency properties of this allocation rule in the face of a set of different noneconomic objectives. They point out

that a proportionally distributed quota tends to produce an additional subsidy to production and a reduced tax on consumption relative to a tariff or standard quota.

Allocation by political decision. In some instances quota allocation is determined on the basis of politics. The U.S. Congress decided the distribution of sugar import quota allocations among sugar exporting countries and the secretary of agriculture sent "certificates of eligibility" to bring sugar into the United States to the governments of the exporting countries. Cuba, which had been the major exporter of sugar to the United States, lost its sugar quota allocation due to disapproval of Fidel Castro's government.

Quota auctions. Australia and New Zealand were two of the very few countries that auctioned quotas to the highest bidders. Auctions provide a market-oriented method of allocating quotas and also raise revenue for the government (Bergsten, Elliott, Schott, and Takacs 1987).

Economic Impact of Import Quotas An import quota is "binding" if the quantity that would be imported without the quota restriction exceeds the quota ceiling. A binding import quota reduces the supply of the product on the domestic market of the importing country and drives up its price. Domestic producers benefit from the quota restriction because the higher price encourages them to produce a larger quantity, which they sell at the higher price. Domestic buyers of the product are worse off because they pay a higher price and buy a smaller amount.

The increase in price of the restricted good in the internal market creates a gap or wedge between the higher internal price and the price of the good on the world market. Traders who can import under the quota system earn a windfall profit, called a "quota rent," because they can buy the product at the going world market price, import, and resell at the higher internal price. The choice of the allocation method for quotas is thus important because it determines who can import and thus who receives the windfall gains from the quota rents.

Import quotas also create net losses in total welfare for the country that imposes them because of resource misallocation and efficiency losses in both production and consumption. On the production side, the quota encourages expansion of domestic production of the restricted good, but the extra production takes place at a higher cost than the price of the good on the world market. The cost of the domestic resources used up to expand domestic production exceeds what it would cost to buy the imported goods that they replace. On the consumption side, buyers lose welfare because they pay a higher price and cut back their purchases of the restricted good in response to the artificially high price. Quotas often cause additional inefficiencies because an overall quota limit on imports of a general product category is often subdivided into smaller quotas for particular varieties of the product, and often further subdivided to specify the country of origin of the imports. For example, the U.S. import quotas on cheese were subdivided into subquotas on different types of cheese (Swiss, provolone, parmesan, cheddar, etc.) and those subquotas were further subdivided to impose upper limits on imports from each exporting country. Thus the quota system arbitrarily determined the distribution of cheese imports by type and source, without regard to cost or consumer preferences (Anderson 1988).

Economic Impact of Export Quotas A binding export quota constrains the amount of a good that can leave the country, forcing more of it onto the domestic market. The larger supply on the domestic market reduces the internal market price. Domestic producers of the restricted good lose because they produce a smaller quantity that sells at a lower price. Domestic buyers or users gain because they obtain the good at a lower price and buy more of it.

Export quotas cause distortions on the production and consumption sides that create overall welfare losses for the economy as a whole. On the production side, artificially constrained exports preclude domestic firms (and thus the economy as a whole) from capturing the gains from selling in the world market at a price above production cost. On the consumption side, the decrease in price encourages domestic buyers to expand purchases, but the extra units they buy could have been sold in the world market at a price higher than the value to domestic users.

Export quotas also generate quota rents because the decrease in the domestic price of the restricted good due to the export quota creates a wedge between domestic and world market price, so whoever can export the good receives a windfall gain equal to the difference between the world market and domestic price. VERs also involve quota rents because the restriction in supply will push up the price of the restricted good in the domestic market of the importing country. Exporters who receive quota allocations to ship to the importing country are likely to receive the quota rents. De Melo and Tarr (1992) calculated very large losses to the United States due to the transfer of quota rents to exporting countries under VER agreements on textiles and apparel, automobiles, and steel.

Comparison of Quotas to Tariffs Quotas influence international trade flows by directly limiting the quantity of a product that can be imported or exported. An import tariff, in contrast, reduces imports through the price mechanism by making imported goods more expensive. A major question in the analysis of international trade barriers is the equivalence of quotas and tariffs (Bhagwati 1968). The question is, "Under what conditions will an import tariff and an import quota that result in the same amount of imports have exactly the same economic impact?" If market conditions remain unchanged, and the product markets are perfectly competitive, an import quota and a tariff that result in the same import quantity will also result in the same domestic price, quantity produced, and quantity consumed. The literature on the equivalence of tariffs and quotas has identified a number of ways in which the impacts of tariffs and quotas differ, however (for detailed explanations see Vousden 1990).

Government revenue. The government collects revenue from an import tariff or an export tax, but import or export quotas generate quota rents, which do not directly go to the government. Governments could in principle attempt to capture what would be quota rents by auctioning quota licenses, but quota auctions are rare (Bergsten, Elliott, Schott, and Takacs 1987). Quota rents may be captured by the nationals of the country imposing the quota: for

example, if import licenses are allocated only to domestic firms importing the good. But if import licenses or permits are given to exporters (as in the case of the U.S. sugar import quotas or under VERs), the quota rents are very likely to be captured by nationals of the exporting country and thus represent an additional loss to the importing country, in addition to the welfare losses from production and consumption distortions. The possibility that quota rents may be lost to the quota-imposing country is one reason to favor tariffs over quotas as a policy instrument to reduce imports or protect a domestic industry.

Noncompetitive market structure. If the domestic industry producing the good in the importing country is not perfectly competitive, the economic impact of a quota and a tariff can differ. If only one domestic firm produces an imported good, that domestic monopoly faces quite different demand conditions for its product under a tariff than under a quota. If the monopoly firm is protected by a tariff, it can sell its product at a price above the world market price, but it cannot control the price at which it can sell. If it attempts to raise its price above the world price plus tariff, buyers can simply buy from foreign suppliers and pay the tariff. The monopoly has no ability to control the domestic price with a tariff. It has no "market power." But if the monopoly is instead protected by an import quota set to allow in the same quantity of imports as the tariff, the firm will raise its selling price because buyers no longer have the option of importing more from abroad than is allowed by the quota. An import quota gives a domestic monopoly "market power" that it would not have under the tariff, which it will exploit by reducing output and charging a higher price. This extra distortion that arises from a quota when domestic industries are highly concentrated is another reason to favor tariffs over quotas.

Changes in market conditions. A tariff and a quota also lead to different outcomes if market conditions change after the tariff or the quota is imposed. The difference arises because a tariff allows imports to increase if market conditions change, whereas a quota does not. Assume a small country that is too small to influence world prices by the

amount that it imports. If a binding quota restricts imports to the quota ceiling and domestic demand for the product subsequently increases, imports cannot increase to satisfy the extra demand, so the domestic price of the product will rise and domestic production will increase. In contrast, if a tariff had restricted imports to the same initial level (resulting in the same initial domestic price, production, and consumption) an increase in domestic demand simply would increase the quantity imported, without any increase in the price, which would remain equal to the world market price plus the tariff.

A tariff and an import quota also lead to different outcomes if international market conditions change. If the world market price of an imported good falls, but domestic demand and supply conditions remain unchanged, the import quota prevents imports from increasing even though their price is falling. The external price decline will not be transmitted into the domestic economy. If a tariff had been used to control imports, falling world prices would make imports less expensive, and the quantity of imports would increase. The world market price decline would be transmitted into the domestic economy. This difference between tariffs and quotas is the main reason governments used quotas to control imports of products subject to domestic price supports.

Upgrading of product quality. If a product category contains varieties of different qualities that sell at different prices, the composition of imports will differ under a tariff regime and a quota regime. Import quotas result in "upgrading," or a shift toward imports of higher-quality, higher-priced varieties, rather than lower-quality, lower-priced varieties. The quota creates a wedge, or price gap, between the domestic price and the world price of the imported products. The given price gap represents a larger percentage price increase for lower-priced goods than for higher-priced goods, so lower-priced goods become more expensive relative to higher-price goods and there is a tendency for buyers to substitute higher-quality goods for lower-quality goods.

Directly unproductive activities. The quota rents earned by those individuals or firms who receive the rights to import under quota provide a powerful incentive to engage in activities to influence quota allocation decisions. Thus quota regimes encourage individuals to spend time and use up real resources in their attempt to influence quota allocation decisions in their favor (Krueger 1974). Tariff revenue accrues directly to the government, so a tariff regime does not promote similar unproductive activities.

Quotas have long been a feature of global trade policies. While the role of quotas has diminished somewhat with recent progress in liberalizing trade in the textile, clothing, and agricultural sectors under the auspices of the World Trade Organization, the analysis of quotas remains an important area in theoretical and applied trade policy research. As such, it informs a preference in the trade policy community for tariffs as a more appropriate trade policy intervention.

See also nontariff measures; tariff rate quotas; tariffs

FURTHER READING

Anderson, James E. 1988. *The Relative Inefficiency of Quotas.* Cambridge, MA: MIT Press. Examines the inefficiencies caused by quotas, including estimates of the cost of U.S. cheese quotas and analysis of quotas under imperfect competition and uncertainty.

Bergsten, C. Fred, Kimberley A. Elliott, Jeffrey Schott, and Wendy E. Takacs. 1987. *Auction Quotas and U.S. Trade Policy.* Washington, DC: Institute for International Economics, 1987. A discussion of auctioning import quotas, with a description of the experience of Australia and New Zealand with their quota auctions.

Bhagwati, Jagdish. 1968. "More on the Equivalence of Tariffs and Quotas." *American Economic Review* 58 (1): 142 46. An analysis of the conditions under which tariffs and quotas have an equivalent economic impact.

Bohi, Douglas R., and Milton Russell. 1978. *Limiting Oil Imports: An Economic History and Analysis.* Baltimore: Johns Hopkins University Press. A presentation of the details of the history and administration of the U.S. petroleum import quota system, with an economic analysis.

Corden, W. M. 1971. *The Theory of Protection.* Oxford: Oxford University Press. An accessible discussion of trade policy instruments that contains an extensive analysis of quotas.

de Melo, Jaime, and David Tarr. 1992. *A General Equili brium Analysis of U.S. Foreign Trade Policy.* Cambridge, MA: MIT Press. An explanation of general equilibrium modeling with estimates of the impact of voluntary ex port restraints on shipments of textiles and clothing, automobiles, and steel to the United States under a va riety of assumptions.

Hamilton, Carl. 1988. "ASEAN Systems for Allocation of Export Licenses under VERs." In *The Political Economy of Manufacturing Protection: Experiences of ASEAN and Australia*, edited by Christopher Findlay and Ross Garnaut. London: Allen and Unwin, 235 47. Reviews alternative methods governments used to decide on the allocation of export quotas among firms.

Krueger, Anne O. 1974. "The Political Economy of the Rent Seeking Society." *American Economic Review* 64 (3): 291 303. The original analysis of directly un productive activities associated with quota regimes.

McCulloch, Rachel, and Harry G. Johnson. 1973. "A Note on Proportionally Distributed Quotas." *American Eco nomic Review* 63 (4): 726 32. An analysis of the eco nomic impact of quota allocation to domestic producers in proportion to their production levels.

Vousden, Neil. 1990. *The Economics of Trade Protection* Cambridge, MA: Cambridge University Press. Contains extensive and detailed analyses of the economic impact of quotas, including losses to consumers, transfers to producers and recipients of quota rents, and the efficiency losses from distortions in production and consumption. Also discusses the effects of quotas in the presence of retaliation by trading partners, the rent seeking costs of tariffs and quotas, product upgrading when quotas are imposed on differentiated products, and a discussion of the comparison of quotas with tariffs when the protected sector is a monopoly.

WENDY TAKACS

real exchange rate

The real exchange rate, a linchpin of international economics, tells us about the overall costs in one country compared with another. Defined as the cost of a bundle of goods in one country relative to the cost of the same bundle of goods in another country, it is conventionally written as the nominal exchange rate adjusted for international differences in prices levels:

$$Real\ Exchange\ Rate \equiv$$
$$Nominal\ Exchange\ Rate \times \frac{Domestic\ Prices}{Foreign\ Prices},$$

where the nominal exchange rate is specified here in terms of foreign currency units per unit of domestic currency.

From this expression, one sees that if the nominal exchange rate fluctuates widely while the price indexes are relatively stable, then the real exchange rate will move with the nominal exchange rate. This indeed has been the pattern under floating exchange rates: real exchange rates have been nearly as volatile as nominal exchange rates have been. Moreover changes in the real exchange rate seem to persist for a very long time. Explaining the behavior of real exchange rates therefore requires both an understanding of why nominal exchange rates are so volatile under floating exchange rate arrangements and an understanding of what keeps prices from fluctuating proportionately.

Theoretical Approaches Most theoretical approaches to understanding real exchange rates focus on the price of goods. These explanations generally fall into two categories. The first category emphasizes the importance of international borders, particularly those aspects of borders that prevent the law of one price from holding internationally. The second category emphasizes what goes on within countries, especially what happens to the relative prices of traded and nontraded goods within an individual country. International failures in the law of one price and international differences in relative prices within countries are related, but they each capture different ideas of what is important in determining real exchange rates.

Across international borders, the law of one price fails for many goods, and its failure is particularly notable for final goods. Its failure can occur most obviously because of transportation costs and official trade barriers, but it also can occur because of noncompetitive market structures, in which prices resist change. In many new open economy macroeconomic models, firms have monopoly power and prices are set as a markup over marginal cost. In these models, the law of one price can fail, at least in the short term. Failures of the law of one price represent impediments to the adjustment of the real exchange rate.

Looking within countries, explanations of real exchange rate behavior emphasize distinctions between different sectors of the economy, such as the traded and nontraded sectors, or the intermediate and final sectors. These explanations tend to rely either on differential changes in productivity across the sectors or on changes in the relative demands across the sectors. All such changes affect the relative prices and the composition of what is produced and consumed in each country. Such changes, in turn, affect

the real exchange rate via the price indicators that we see in its definition. The most prominent of these explanations is the Balassa-Samuelson effect.

The other explanations of real exchange rate behavior are primarily shorter-term explanations. Changes in preferences across an economy's traded and nontraded sectors whether due to changes in consumption or government expenditures will affect relative prices for as long as it takes for the factors of production to move across the sectors within a country. (Some of the new open economy models fit into this category as well.) In the short run, the supply of nontraded goods is inelastic, and the demand for nontraded goods matters. So changes in the consumption of nontraded goods will affect relative prices, and real exchange rates will change.

Empirical Challenges Despite much work, little is known about the empirical determinants of real exchange rates. A large number of econometric studies have attempted to determine whether the exchange rate converges over time to an equilibrium value that is, whether it is stationary. The focus on this particular issue arises in large part because it is equivalent to knowing whether relative purchasing power holds, in which case nominal exchange rate changes and inflation differences eventually offset each other. The persistence of real exchange rate changes, however, means that tests even of this fairly clear-cut hypothesis lack statistical power, even with long time series or with panel data.

Econometric work that attempts more broadly to discern the real exchange rate's most important determinants and thereby to distinguish among competing theories has had only limited success. The real exchange rate's volatility and persistence are not matched by observations of its theoretical fundamentals, and it is difficult to generate endogenous persistence in real exchange rate models without introducing other counterfactual implications.

Thus empirical studies have been able to tell us only a few things. First, real exchange rates tend to move with nominal exchange rates, and their behavior correspondingly depends on prevailing nominal exchange rate arrangements. Second, real exchange rates are volatile. Finally, their changes are persistent.

Even this last conclusion is subject to the criticism that it may be an artifact of poor measurement.

Measuring the Real Exchange Rate Empirical work on real exchange rates involves many judgments about how to construct real exchange rate measures. Most important, appropriate price indicators and suitable aggregation methods must be selected. Typically, price indexes, such as the consumer price index or the producer price index, are used. Just as those indexes show price changes rather than price levels, real exchange rates so constructed will gauge real exchange rate changes, rather than their levels. These measures tell us what the real appreciation or depreciation has been relative to some point in time, but they do not provide the real exchange rate's actual value. Calculation of the real exchange rate's level requires information about price levels, information that is not always readily available. Such distinctions between *indexes* and *levels* can be important in debates about exchange rate policies and trade policies. For example, an index can tell how much the value of the Chinese currency has changed since the Asian financial crisis in 1997 98, but only a measure of the real exchange rate's level can tell us if it is weak or strong relative to another currency or relative to a theoretical value.

One also faces the question of what categories of prices to use. Consumer prices are used most often. Producer and wholesale price indexes, and even export price indexes, are sometimes chosen over the consumer price index because they better represent the notion of export competitiveness, which often is of particular interest. As the set of prices narrows increasingly toward including only actually traded goods, however, it also becomes increasingly like a measure of the terms of trade the relative price of imports and exports and less a measure of the real exchange rate.

Finally, since currency values so often diverge, real exchange rate measures can be sensitive to the set of countries included and to their weights. A country's real exchange rate often is measured bilaterally, that is, against only one other country. A broader multilateral measure can differ markedly from a bilateral one, however. For example, a country that

pegs its nominal exchange rate against that of a single trading partner is likely to have a relatively stable bilateral real exchange rate against that country. Against other countries, however, its real exchange rate could be very unstable, either rising or falling. A multilateral or effective real exchange rate will capture those broader movements.

Whether one is interested in questions about the fundamental determinants of the real exchange rate or has questions about the experience or policies of a single country, all of the measurement choices affect the measures and thereby shape the conclusions.

See also Balassa-Samuelson effect; currency crisis; effective exchange rate; equilibrium exchange rate; exchange rate regimes; exchange rate volatility; New Open Economy Macroeconomics; purchasing power parity

FURTHER READING

Cheung, Yin Wong, Menzie Chinn, and Antonio Garcia. 2005. "Empirical Exchange Rate Models of the 1990s: Are Any Fit to Survive?" *Journal of International Money and Finance* 24:7, 1150 75. Updates Meese and Rogoff (1983) and finds models to have limited and uneven empirical success.

Froot, Kenneth, and Kenneth Rogoff. 1995. "Perspectives on PPP and Long Run Real Exchange Rates." In *Handbook of International Economics*, vol. 3, edited by Gene Grossman and Kenneth Rogoff. Amsterdam: Elsevier Science, 1647 88. Reviews empirical studies of real exchange rates, emphasizing tests of stationary and the Balassa Samuelson effect.

Meese, Richard, and Kenneth Rogoff. 1983. "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics* 14: 3 24. The classic real exchange rate horserace, in which a random walk beats all.

Obstfeld, Maurice, and Kenneth Rogoff. 1995. "Exchange Rate Dynamics Redux." *Journal of Political Economy* 103: 624 60. A pioneering piece in new open macro economics: combines dynamic optimization with im perfect competition and nominal rigidities in a single theoretical approach to modeling exchange rates and current accounts.

HELEN POPPER

regional currency areas

See common currency; optimum currency area (OCA) theory

regional development banks

Regional development banks (RDBs) play a critical role in the international financial architecture. They provide development finance to their poorer member countries on highly concessional terms, mobilizing between U.S. $15 billion and $20 billion a year, compared with about U.S. $20 billion for the World Bank. By comparison, foreign direct investment flows represent more than U.S. $300 billion per year, albeit concentrated in a few emerging markets.

RDBs are central components of the constellation of the 20 or so multilateral development banks that emerged after World War II: the World Bank; five RDBs, four dealing with transition and developing countries (the African Development Bank, the Asian Development Bank, the European Bank for Reconstruction and Development, and the Inter-American Development Bank) and one, the European Investment Bank, dealing mainly with members of the European Union, although the level of its operations in developing countries is significant at around U.S. $3 billion a year; and 13 subregional development banks of varying size, such as the Council of Europe Development Bank, the Caribbean Development Bank, and the West African Development Bank.

Lending for Development The RDBs' mandate has evolved over time, as has the rationale for development finance. They are multilateral organizations mobilizing development finance to their member countries to promote growth, encourage development, and reduce poverty. They act as financial intermediaries, leveraging resources from international capital markets that they lend-on at advantageous conditions to their members, which often have limited and volatile access to those markets. Each of them has built up a corpus of economic and technical expertise, and in 2004, their combined outstanding loans were more than U.S. $1 trillion, of which U.S.

$128 billion was for the RDBs operating in transitional and developing countries (compared with U.S. $225 billion for the World Bank).

Two general characteristics distinguish RDBs from private financial institutions and bilateral donors: their multilateral shareholding structure and preferred creditor status, and a subsidized capital base and access to other subsidies. Their financing model is similar to that of the World Bank, with which there is a certain degree of competition. They are perceived as generally less efficient and more politicized, imposing less stringent conditions than the World Bank. Their comparative advantage resides in their regional constituency, governance structures, and closeness to their members, as well as the low interest rates they charge their borrowing members.

Features The financing model of RDBs consists of three main features: a financial leverage mechanism, a preferred creditor status, and a soft-lending window. RDBs borrow on international capital markets, backed by a capital structure with a low paid-in amount (direct cash contributions representing between 4 and 20 percent of subscribed capital, or issued share capital). They are thus able to borrow in international capital markets at lower rates than their borrowing members could individually, and the RDBs pass on to their members the low interest rates they pay. Their preferred creditor status is anchored in the support of their shareholding governments, reflected in capital contributions and pledges to soft-lending windows providing loans with extended grace periods, longer amortization and lower interest rates than conventional bank loans, and the incentives for borrowing countries to service their debt in order to secure continued access to concessional sources of development finance. Their membership typically includes regional borrowing members, as well as regional and nonregional nonborrowing countries.

Members contribute to the authorized capital, partly paid-in but mainly in the form of callable capital should an RDB become unable to meet its obligations to bondholders. Each member's financial contribution is a function of its relative affluence and determines its voting rights as a shareholder. Nonborrowing members contribute grant resources to their soft-lending windows, which they replenish every three to five years, allowing them to lend-on at highly concessional rates and on a very long-term basis to poorer members.

Functions RDBs provide nonconcessional loans and concessional credits, as well as nonrepayable grants, technical assistance, and policy advice. They also generate and disseminate knowledge, through advisory and analytical work, which shapes public policies and informs policy reform. Typically, they have four main lending instruments: a window providing nonconcessional loans to its wealthier members, financed with ordinary capital and with a repayment period typically of 15 30 years; a soft-lending window established as a separate development fund providing long-term concessional credits to poorer members with low creditworthiness, essentially in Africa and Asia, with a large grant element and a repayment period of up to 40 50 years; a private sector lending window operating on a commercial basis and providing a mix of financing instruments (loans, equity finance, risk management, intermediary finance); and a guarantee window promoting foreign direct investment.

Instruments RDBs provide two main types of sovereign loans or credits. Policy-based loans (also called adjustment or programmatic loans) provide general budget support against a program of reform and are linked to disbursement conditions, at times earmarked to specific sectors such as health, education, or social protection. Investment loans are designed to support specific development projects in sectors such as water, sanitation, infrastructure, or energy utilities. The banks increasingly support subnational governments and large municipalities, sometimes without sovereign guarantee from the central government. They are also prompted to increase their support to the private sector in better-off regions, especially small and medium enterprises, directly or through commercial banks (lines of credit).

Stepping Up to the Future The RDBs, like most multilateral development banks, are confronted with multifaceted challenges from a changing environment and the evolving financing needs of their members, especially middle-income countries. The

rapid development of international capital markets and recent financial crises have prompted many reassessments of the role of multilateral finance, including by the G7 since 2001. Key challenges evolve around the RDBs' corporate mandate, comparative advantage, and lending strategies.

Corporate Mandate Many attempts are being made to redefine the banks' mission, as part of a gradual rethinking of the purpose and effectiveness of multilateral lending. RDBs are to focus more narrowly on reducing poverty and inequality and address a broader set of development challenges, including governance and anticorruption strategies. They are pressed by their richer shareholders to deliver results, and by civil society to show greater corporate social responsibility in their sovereign loans and equity investments in sensitive sectors such as extractive industries. They are under external pressure to increase transparency and accountability in their internal governance and modus operandi and are often criticized for operating as foreign policy extensions of powerful shareholders, especially the United States.

Strategic Selectivity RDBs are pressed to act more strategically and selectively, concentrating in those countries with sound institutions and good governance where aid has been found to be more effective. Calls for more selective engagement are part of the broader debate on the balance between grants and loans, lending instruments and pricing policies. Shifting to grant aid, it is argued, would help avoid the reaccumulation of unsustainable levels of debt. A

Table 1
Basic facts about regional development banks, 2005

	World Bank	Inter American Development Bank	African Development Bank	Asian Development Bank	European Bank for Reconstruction and Development[b]
Year established	1945	1959	1964	1966	1990
Membership	184	46	77	63	62
Soft window	IDA	FSO	AfDF	AsDF, NTF	
Private sector window	IFC	IIC, MIF			
Credit rating	AAA	AAA	AAA/AA+ (2003)	AAA	AAA
Subscribed capital (2004, U.S.$bn)	178	101	32	32	32
Of which paid in	*11.5*	*4.34*	*3.24*	*2.32*	*6.4*
Total loan portfolio[a] (disbursements 2005, U.S.$bn)	225.3	55.0	33.8	24.5	45.5
Lending volume (gross disbursements 2005, U.S.$bn)	18.7	5.3	1.8	4.7	2.8
Net lending (net disbursements 2005, U.S.$bn)	2.2	0.2	0.7	0.4	0.3
Total voting rights for nonconcessional window for (%):					
Borrowing countries (2006)	38.2	50.4	56.5	40.5	11.7
Nonborrowing countries (2006)	61.8	49.6	43.5	59.5	88.3

Sources: Annual reports of the World Bank and RDBs.
[a] Measured as total loans outstanding and excluding private sector affiliates.
[b] Exchange rate used for EBRD: €1 = U.S.$1.28.

major debt relief initiative for highly indebted poor countries was launched in 1996 (enhanced in 1999), amounting to about U.S. $100 billion in nominal terms. In 2005, the Multilateral Debt Relief Initiative agreed to cancel the eligible debt of 22 poor countries, in particular in Africa, owed to the concessional lending arms of the African Development Bank and Inter-American Development Bank (as well as the World Bank and the International Monetary Fund).

In addition to pressuring the RDBs for greater reliance on grants in low-income countries with limited fiscal capacity, observers often question the banks' relevance and effectiveness in middle-income countries, partly as a result of developing countries' increased access to capital markets. In 2000, the U.S. Congress's Meltzer Commission recommended that the RDBs essentially end nonconcessional lending to middle-income countries and that their assistance to low-income countries be mainly in the form of grants. The commission also suggested that RDBs should assume most of the lending to their respective regions, except for Africa. Nevertheless, the emerging markets crises of the late 1990s showed that middle-income countries' access to private capital remains volatile and precarious.

Changing Context In a changing environment, the RDBs' importance as a source of development finance is declining, overtaken by private sector flows and remittances. Moreover grant financing from bilateral donors is set to increase sharply. New actors, especially China, are emerging as alternative sources of development finance, with fewer strings attached. In addition, the RDBs' loan portfolios are maturing, resulting in transfers out of developing countries as repayments and prepayments are made. In 2005, for example, net lending was nil, if not negative, reflecting patterns of "defensive lending," whereby new loans are made to ensure that past loans are repaid (see table 1).

Repositioning Efforts The RDBs are thus seeking to reposition themselves along different directions, through a more selective approach in middle-income countries, a sharper poverty focus in lower-income countries, and greater synergies with the private financial sector. They are concentrating on good performers with sound economic policies in order to improve portfolio performance, tackling global public goods, and restricting their balance of payment support, especially emergency lending in crisis situations. Borrowers see competition among the multilateral development banks as providing them with greater choice and therefore improving loan quality, but specialization and cooperation between the World Bank and regional and subregional development banks are areas for further improvement.

See also aid, international; World Bank

FURTHER READING

Bergsten, Fred. 2000. *Reforming the Multilateral Development Banks.* Washington, DC: U.S. Senate. Testimony before the Subcommittee on International Trade and Finance Committee on Banking, Housing, and Urban Affairs, 8 June.

Bezanson, Keith, Francisco Sagasti, Silvia Charpentier, and Ricardo Gottschalk. 2000. *A Foresight and Policy Study of the Multilateral Development Banks.* Stockholm, Sweden: Ministry of Foreign Affairs. The report looks at the current pressures on the MDBs to reform their governance and retarget their operations in terms of their triple function of mobilizing financial resources; building institutional capacity and brokering knowledge; and finally, providing regional and global public goods.

Culpeper, Roy. 1997. *Titans or Behemoths? The Multilateral Development Banks.* London: Intermediate Technology. This final volume in the five title series on the MDBs examines the effectiveness of the MDBs and studies the overarching role of the World Bank. Arguing for a process of reform of the global governance system rather than its reinvention under a new framework, Culpeper urges the MDBs to build on their efforts to improve transparency and accountability.

Gurría, José Angel, and Paul Volcker. 2002. *The Role of Multilateral Development Banks in Emerging Market Economies.* Washington, DC: CEIP. This report from a commission cochaired by former Mexican finance minister José Angel Gurria and former U.S. Federal Reserve chairman Paul Volcker, assesses the activities of the MDBs in emerging market economies and recom

mends changes on long standing MDB approaches to ward these countries.

International Financial Institutions Advisory Commission. 2000. *Report of the International Financial Institutions Advisory Commission.* Washington, DC: U.S. Congress. The International Financial Institution Advisory Com mission, also known as the Meltzer Commission named for its chair, professor Allan Meltzer was es tablished by the U.S. Congress in November 1998 "to recommend future US policy toward several multilateral institutions: the IMF, the World Bank group, the re gional development banks." The commission's influ ential majority report proposed changes to the opera tions of the IMF and especially to those of development banks such as the World Bank, which the majority re commended should withdraw from lending to so called middle income countries.

Rodrik, Dani. 1995. "Why Is There Multilateral Lending?" NBER Working Paper No. 5160. Cambridge, MA: National Bureau of Economic Research. Reviews the rational for multilateral lending to developing countries.

CARLOS SANTISO AND EMILE ROBERT PERRIN

■ regionalism

Enshrined in Article I of the World Trade Organization's (WTO's) General Agreement on Trade and Tariffs (GATT) is the key principle of the multilateral trade system the principle of nondiscrimination which prevents member countries from discriminating against imports based on the country of origin. In an important exception to this central prescript, however, the GATT (Article XXIV) permits WTO members to enter into preferential trade agreements (PTAs), provided these preferences are complete. In so doing, it sanctions the formation of free trade areas (FTAs), whose members are obligated to eliminate internal import barriers, and customs unions (CUs), whose members additionally agree on a common external tariff against imports from nonmembers. In addition, the Enabling Clause allows tariff preferences to be granted to developing countries (in accordance with the Generalized System of Preferences) and permits preferential trade agreements among developing countries in goods trade. Among the more prominent existing PTAs are the North American Free Trade Agreement (NAFTA), the European Economic Community (EEC), and the European Free Trade Association (EFTA), all formed under Article XXIV, and Mercosur (the CU encompassing Argentina, Brazil, Paraguay, and Uruguay) and the ASEAN (the Association of Southeast Asian Nations) Free Trade Area (AFTA), both formed under the Enabling Clause. Other forms of integration, such as the formation of a common market, involve the further elimination of barriers to factor mobility, as in the Common Market for Eastern and Southern Africa (COMESA) and the use of a single currency to form a common currency area (as in the European Union). Since preferential integration has often (but not always) involved geographically proximate countries, the term *regionalism* has been used to describe such regional groupings and sometimes to describe the broader phenomenon of preferential integration itself.

Economic Analysis Motivated by ongoing discussions concerning optimal trade arrangements in the postwar period, especially over the possibility of a European customs union, Viner (1950, 41 50) developed a seminal analysis of the economics of preferential trade. Viner's analysis disputes the presumption that cutting tariffs necessarily improves welfare. On the one hand, because of discriminatory liberalization, there will be commodities that a member country may "newly import from the other but which it formerly did not import at all because the price of the protected domestic good was lower than the price of any foreign source plus the duty." Viner calls this shift from a high to a lower cost point "trade creation" and associates it with welfare improvement for the importing country. He also argues that, on the other hand, "there may be other commodities, which one of the members will now newly import from the other," whereas before the PTA it "imported them from a third country, because that was the cheapest possible source of supply even after the payment of duty." He calls this shift in imports from a low-cost third country to a higher-cost member country

"trade diversion," associating it with an increase in the cost of imports and, thus, welfare losses for the importing country.

The demonstration that preferential trade liberalization may be welfare decreasing stimulated a substantial theoretical literature on the "static" welfare effects of PTAs. Post-Vinerian analysis of the welfare effects of preferential trade developed examples of both welfare-improving trade diversion and welfare-decreasing trade creation in general equilibrium contexts broader than those considered by Viner (see Panagariya 2000 for a comprehensive discussion). The intuitive appeal of the concepts of trade creation and trade diversion and supporting empirical findings, however, have ensured their continued use in the economic analysis of preferential trade agreements, especially in policy analysis.

It has sometimes been argued that countries entering into preferential arrangements with geographically proximate countries are likely to do better than in agreements with distant countries, because the former are more likely than the latter to be trade creating. The theoretical literature has provided a number of examples, however, in which, between two otherwise identical potential partners, a country achieves a superior outcome by granting trade preferences to the distant partner (see Panagariya 2000). Moreover empirical analysis of this question has not found any support for the hypothesis that regional trading partners should be considered "natural" partners in the context of preferential trade (see Krishna 2005).

The generally ambiguous welfare results provided by the theoretical literature raised an important question relating to the *design* of necessarily welfare-improving PTAs. A classic result by Kemp and Wan (1976) provided a welfare-improving solution for the case of CUs. Starting from a situation with an arbitrary structure of trade barriers, if two or more countries freeze their net external trade with the rest of the world through a set of common external tariffs and eliminate the barriers to internal trade (which implies the formation of a CU), the welfare of the union as a whole necessarily improves and that of the rest of the world does not fall. A Pareto-improving

CU is thus achieved. Panagariya and Krishna (2002) have provided a corresponding construction of necessarily welfare-improving FTAs where, in analogy with Kemp and Wan, external trade is frozen for each FTA member country. Common to these analyses of welfare-improving preferential trade blocs is the elimination of trade diversion achieved in each case by freezing external trade. To ensure this outcome, the trade barriers imposed on nonmembers must generally be less restrictive than before. Since existing GATT rules require only that external barriers not be raised, some trade diversion is likely in practice.

Recent analysis in the literature has focused on issues concerning the effects of preferential agreements on the multilateral trade system. Will PTA membership encourage governments to expand their agreements to include new countries or will there instead be political-economy incentives to keep new countries out? (Or, as Bhagwati 1993 has phrased it, will trade blocs serve as "building blocs" or "stumbling blocs" in the path to multilateral free trade?) Does the answer depend on the (trade-creating or trade-diverting) nature of the PTAs involved?

Modeling the endogenous determination of trade policy while emphasizing the role of powerful domestic interests in influencing policy, it has been argued that PTAs that divert trade are more likely to win internal political support. This is so because the political costs of liberalization are alleviated when gains to domestic firms of member countries come largely at the expense of outsiders. It follows that such PTAs will lower the incentives for any subsequent multilateral liberalization due to the fact that producers in trade-diverting PTAs may oppose multilateral reform since this would take away the gains from benefits of preferential access that they enjoyed in the PTA that diverted trade to them. Indeed, the incentives for further multilateral liberalization may be completely eliminated. The literature has obtained similar results when trade policy is determined by majority voting. If bilateral agreements occur between similar countries, so that there are gains from trade but no major income distributional shifts, bilateral agreements could render multilateral liberalization infeasible, as the latter may involve income
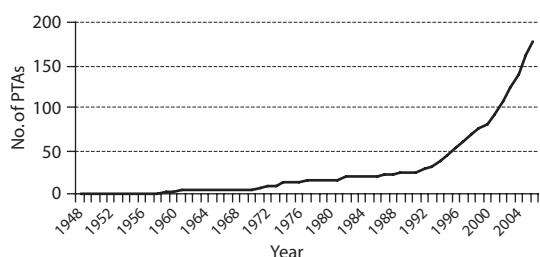
Figure 1

PTAs in force by date of entry into force. Source: World Trade Organization.

distributional changes that worsen the lot of the median voter. These analyses demonstrate the possibility that preferences in trade may have a negative impact on multilateral liberalization incentives for member countries. On the other hand, nonmember countries may see their incentives for multilateral liberalization increased. Thus, it has been argued that PTA expansion could have a "domino" effect increasing the size of a bloc increases the incentive for others to join it (as they then gain preferential access to increasingly large markets). Assuming open membership rules (i.e., insiders do not oppose the entry of new members who abide by the same rules as the members), it has been shown that the successive expansion of a PTA could then lead to multilateral free trade, but this desired outcome is not achieved under alternative membership rules (see Krishna 2005 for a survey discussion).

A half century of research has significantly advanced our understanding of the implications of trade discrimination even if the frequently equivocal theoretical and empirical results have established among economists and policymakers an ambivalent attitude toward preferential trade agreements. Concerns regarding the fragmentation of the world trade system, however, have grown with the rapid proliferation of preferential trade in recent years (see figure 1). Several hundred PTAs were in existence in 2007 (with many countries belonging to multiple PTAs), and several more were in process. With this inexorable erosion of nondiscriminatory disciplines within the trade system, research on preferential trade is certain to remain central to the field of international trade policy for many years to come.

See also common currency; common market; customs unions; free trade area; multilateralism; nondiscrimination

FURTHER READING

Bhagwati, J. 1993. "Regionalism and Multilateralism: An Overview." In *New Dimensions in Regional Integration*, edited by Jaime de Melo and Arvind Panagariya. Cambridge: Cambridge University Press, 22 46. An insightful historical and analytical survey of preferential trade and its interaction with the multilateral trade system.

Kemp, M., and H. Wan. 1976. "An Elementary Proposition Concerning the Formation of Customs Unions." *Journal of International Economics* 6 (1): 95 97. A seminal contribution describing the design of necessarily welfare improving customs unions.

Krishna, P. 2005. *Trade Blocs: Economics and Politics*. Cambridge: Cambridge University Press. A discussion of recent developments in the economic, political, and quantitative analyses of preferential trade agreements.

Panagariya, A. 2000. "Preferential Trade Liberalization: The Traditional Theory and New Developments." *Journal of Economic Literature* 38 (2): 287 331. A comprehensive survey of the theoretical literature on preferential trade agreements.

Panagariya, A., and P. Krishna. 2002. "On the Existence of Necessarily Welfare Improving Free Trade Areas." *Journal of International Economics* 57 (2): 353 67. A theoretical contribution describing the design of welfare improving free trade areas.

Viner, J. 1950. *The Customs Unions Issue*. New York: Carnegie Endowment for International Peace. A seminal contribution characterizing the economic consequences of preferential trade.

PRAVIN KRISHNA

■ remittances

International remittances, the sums of money and goods that immigrants send home, have recently captured the attention of bankers, economists, and policymakers. Previously, little effort was expended
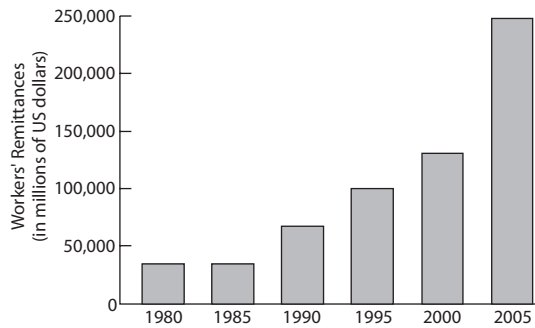
Figure 1

World receipts of workers' remittances. Source: World Bank, World Development Indicators.

to measure and analyze these flows because they were thought to be small in magnitude and of little significance for most countries. Evidence to the contrary has motivated policymakers and others to pay closer attention to the measurement, determinants, and impact of remittances.

Trends and Geography Using a broad definition of international remittances — workers' remittances and compensation of employees — we find that these flows increased from U.S. $37 billion in 1980 to U.S. $249 billion in 2005 (World Bank Indicators). This almost sevenfold increase (see figure 1) exceeds the growth of many macroeconomic variables, including world trade. Although some of the growth in remittances is simply due to better tracking and reporting of these flows at the aggregate level, microeconomic surveys also indicate that there has been brisk growth in the flows of resources from immigrants to their families. It is likely that this growth will persist as an important by-product of world migration.

Figure 2 reports on the worldwide distribution of remittance receipts, revealing that 50 percent of total receipts flow to Latin American, East Asia, and the Pacific regions. South Asia, the Middle East, and North Africa account for about another quarter of flows, while 14 percent flows to Europe and Central Asia. Sub-Saharan Africa, the region most in need of resource inflows, has the smallest share — only 4 percent.

Data Sources Individuals interested in obtaining yearly country-level data on international remittances can consult *The Balance of Payments Statistics Yearbook*, a publication of the International Monetary Fund (IMF). Since these data originate from reports submitted by member nations, more detailed information on remittances can sometimes be obtained directly from reporting agencies of individual countries. For example, Mexico provides monthly data on remittances on its Web site (www.banxico .org.mx). These are disaggregated by transmission method — whether the receipts arrived by money order, check, electronic transfer, or cash. In contrast, understanding how remittances affect household behavior (e.g., schooling, labor force participation) requires obtaining microeconomic data that report on both the household receipt of remittances and household characteristics. Many national income and expenditure surveys, a handful of censuses, and some specialized migration surveys provide this information.

Definition and Measurement of International Remittances At the time of this writing, the term *remittances* is used to refer to one or a combination of three separate concepts: workers' remittances, compensation of employees, and migrants' transfers. The first of these, workers' remittances, are current transfers by nonresidents to resident households, for example, the sums of money that a Dominican national with *U.S. residency* sends to her family in the Dominican Republic each month. In contrast, if the Dominican had been hired to work in the United States for five months only, her earnings net of travel expenses would be classified as compensation of employees. The main distinction in this latter case is her classification as a *Dominican resident* due to the temporary (less than one year) nature of her stay in the United States. The third category, migrants' transfers, are the changes in asset designations that take place when an individual changes residency. Take the case of an individual moving from the United States to Thailand. His U.S. saving account changes from being owned by a *U.S. resident* to being owned by a *Thai resident* and would be recorded under migrants' transfers. In practice there is considerable confusion and practical difficulties regard-
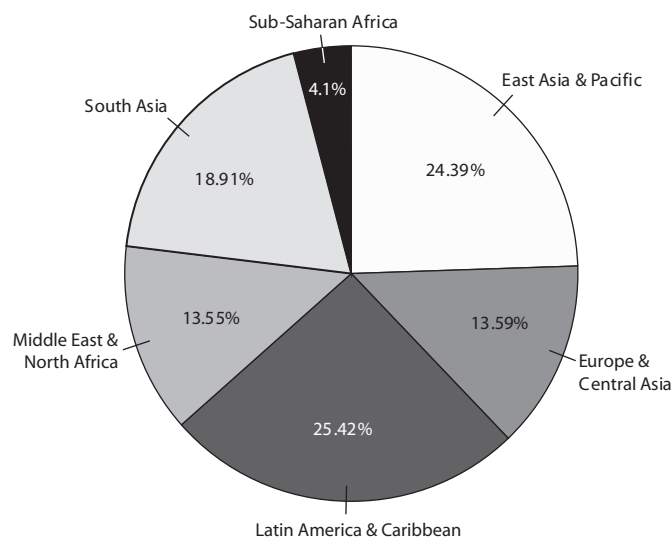
Figure 2

Share of world remittance receipts. Source: World Bank, World Development Indicators.

ing the three designations of remittances. Efforts to change and simplify the reporting and categorization of these international money flows were under discussion as of 2005 (see International Monetary Fund 2005).

Aside from these definitional difficulties, the researcher is likely to be challenged by the noncomparability of data across countries and over time. Although annual data on remittances are derived from standardized reports submitted to the IMF, countries still use different methodologies to collect the data. Some countries measure remittances by aggregating reports of inflows by the banking system and by money transfer firms such as Western Union. Different country-specific reporting requirements and thresholds make for variations in the scope of these reports, however. In other cases, central banks estimate flows using household surveys. Reliability, in these cases, rests with the ability of the survey instruments to capture all flows. Furthermore some surveys ask only about monetary remittances while others collect information on both monetary and in-kind transfers.

Nonuniformity of series across time can cause additional problems. There are variations in the abilities of statistical agencies to track different methods of transmission. It may be relatively easy for central banks to account for transfers made through officially regulated channels (e.g., banks and money transfer firms). Informal transfers those that are hand carried, mailed as cash, or transmitted through hawalas are harder to track (a participant in the hawala system provides a sum to a hawaladar who, in turn, directs a counterpart in another country to pay out that same sum of money to the intended recipient). Variations in policies and regulations can affect the choice of transmission method of immigrants remitting money home, which, in turn, affects the officially reported flows but not the actual flow of money across borders. For example, in 2002 a number of large banks chose to recognize the Mexican *matricula consular* (identification card issued by Mexican consulates to Mexicans residing outside of Mexico) as a valid form of identification, thereby facilitating the opening of bank accounts by many unbanked Mexican immigrants in the United States. An observed surge in *recorded* remittances following the policy change was likely due to shifting from harder to track informal transmission methods to more easily measured formal bank channels.

Determinants of Remittances To appreciate the impact of remittances it is helpful to understand why migrants remit. What motivates remittances? Do remitters respond to economic variables in both the host and home communities? Do incentives to remit change with policy? If we can describe their determinants, we can better predict and influence remittances.

Various motives for remitting have been put forth and tested (Rapoport and Docquier 2006). The motive most commonly cited is altruism — to facilitate higher levels of consumption by the family left in the home country. In fact, the prospect of earning higher wages abroad and sharing those with family back home is thought to be at the heart of much migration. But migration may also be motivated by a desire to accumulate assets, to purchase housing, land, or capital goods, to begin or expand a business in the origin community, or to cover for risky ventures away from home.

Many find it useful to think of migration and remittances as an overall strategy to permit the family to gain from the geographic dispersion of its members (Lucas and Stark 1985). Rosenzweig and Stark (1989) provide a good example of this, demonstrating how marriage with migration takes place to diversify income streams. Total household income remains more stable in these geographically dispersed households as members suffering from income shortfalls are compensated with income transfers by those in preferable situations, lowering consumption variance for all.

Isolating the determinants of remittances is helpful for predicting how remittances may respond to policy and to changing economic conditions. If migration and remittances take place to diversify income, remittance flows may rise and fall with economic shocks. If migration takes place to take advantage of higher earnings to purchase large-ticket items, remittances may subside once the purchase is made. If remittances are purely altruistic in nature, we can expect to observe that they flow and ebb with crisis situations in the home community. If remittances are undertaken to invest in the community of origin, local economic conditions and exchange rate policy are apt to have impacts on the volume and timing of flows.

Impacts of Remittances While many argue that remittances are beneficial, raising living standards, others emphasize negative effects of remittance inflows. This has led to confusion regarding appropriate policy. Should remittances be supported and encouraged or should we impede and slow growth in these money transfers? As with many phenomena, the issue is complex with both viewpoints earning currency.

Macroeconomic impacts: In an often-cited report, Ratha (2003) lauds remittances as more stable and reliable than foreign direct investment or foreign aid and thus an important source for financing development. Chami, Fullenkamp, and Jahjah (2005) take issue with this view, finding remittances to be negatively correlated with economic growth, compensatory in nature, and sent to families during hard times. As such they do not view remittances as a source of funding for investment and growth. Fajnzylber and López (2007) find that remittances contribute to surprisingly modest economic growth, while Adams and Page (2005) credit remittances with reducing both the rate and depth of poverty in developing countries.

The scorecard on remittances becomes even more difficult to interpret when we consider that remittances have uneven, time-sensitive, and redistributive effects. For example, given that international migration is costly, the benefits from remittances often bypass the poorest, who lack the means to finance migration. But this is found to change over time because as migration networks become more widespread, individuals lower in the income distribution obtain the means to migrate and in turn benefit from remittances too (McKenzie 2005).

Additional redistributive impacts may result when remittance spending patterns cause the relative prices of nontraded goods to increase, causing real exchange rate appreciation. This places the export sector in a less competitive position and results in an additional set of gainers and losers. If national policy

is set on fueling economic development via export growth, changing real exchange rates due to remittances may thwart those efforts.

Microeconomic impacts: Understanding the microeconomic impacts of remittance inflows might help settle some of the controversy regarding their macroeconomic impacts. Are remittances all "squandered away" on consumption and thus have no chance to contribute toward longer-term economic growth? Or are these flows invested, increasing the physical and human capital stock? Surveys that simply ask senders what the transfers are intended for have sometimes been consulted to determine the share invested and the share consumed. But while senders claim to transfer money for specific purposes, it lies with the recipient to actually allocate the inflows.

Many researchers use microeconomic-level surveys that link remittances with household characteristics. Using such, Cox Edwards and Ureta (2003) found that remittances positively affect the level of schooling in children. Others, however, have found only very modest increases in schooling levels due to remittances. Differences in research outcomes are likely due to difficulties in separating the "migration effect" from the "remittance effect." Remittances are often a by-product of migration by a household member. Although remittances may loosen liquidity constraints, facilitating school attendance, the absence of a family member may force children to assume the role of an absent member in home or market work, reducing time available for school.

What is the impact of remittances on work behavior? Do remittances promote dependency by inducing individuals who would otherwise work to refrain from supplying their labor? Amuedo-Dorantes and Pozo (2006) find that women remittance recipients in Mexico appear to work fewer hours, perhaps because the additional income allows the women to tend to children instead. Men also appear to change their behavior but not in terms of hours of work. They appear to move away from formal sector work and toward informal sector work much of which is self-employment. Perhaps

remittances facilitate self-employment by providing resources to acquire the necessary capital stock or a "backup fund" to permit the recipient to engage in riskier entrepreneurial ventures. Evidence that remittances are used for the acquisition of physical capital is furthered by Woodruff and Zenteno (2007), who find that attachments to migration networks in Mexico are associated with increases in investments in microenterprises.

Since migrants are complex individuals facing a variety of choices, circumstances, and constraints, it stands to reason that there are many different motives for remitting. Differing motives will result in varying impacts of remittances. In many cases the impacts will benefit recipients, but in other circumstances they might not. Nevertheless, remittances will likely continue to grow and take on greater importance. Continued research to further our understanding of remittances will permit the tailoring of policy to better unlock their potential to do good for recipient households and nations.

See also balance of payments; brain drain; brain gain; brain waste; migration governance; migration, international; real exchange rate

FURTHER READING

Adams, Richard H. Jr., and John Page. 2005. "Do Inter national Migration and Remittances Reduce Poverty in Developing Countries?" *World Development* 33 (10): 1645 69. A large study covering the impact of remit tances on poverty rates.

Amuedo Dorantes, Catalina, and Susan Pozo. 2006. "Mi gration, Remittances, and Male and Female Employ ment Patterns." *American Economic Review Papers and Proceedings* 96 (2): 222 26. Covers labor supply impacts of remittances.

Chami, Ralph, Connel Fullenkamp, and Sanir Jahjah. 2005. "Are Immigrant Remittance Flows a Source of Capital for Development?" *IMP Staff Papers* 52 (1): 55 81. Challenges the notion that remittances are stable funding for development.

Cox Edwards, Alejandra, and Manuelita Ureta. 2003. "In ternational Migration, Remittances, and Schooling: Evidence from El Salvador." *Journal of Development*

Economics 72 (2): 429 61. Addresses remittances' im pacts on schooling,

Fajnzylber, Pablo, and J. Humberto López. 2007. *Close to Home: The Development Impact of Remittances in Latin America*, Washington, DC: World Bank. An extensive treatment of remittance flows to Latin America.

International Monetary Fund. 2005. *Remittances and Sta tistics*. Downloadable from http://www.imf.org/external/ np/sta/bop/remitt.htm. Papers and presentations pro duced for a joint World Bank/IMF meeting to discuss the measurement of remittances.

Lucas, Robert E. B., and Oded Stark. 1985. "Motivations to Remit: Evidence from Botswana." *The Journal of Poli tical Economy* 93 (5): 901 18. Pioneering study of the economics of remittances.

McKenzie, David J. 2005. "Beyond Remittances: The Ef fects of Migration on Mexican Households." In *Inter national Migration, Remittances, and the Brain Drain*, edited by Çaglar Özden and Maurice Schiff. Wash ington, DC: World Bank, 123 47. Highlights the dis tinct impacts of migration and remittances on household behavior in Mexico.

Rapoport, Hillel, and Frederic Docquier. 2006. "The Economics of Migrants' Remittances." In *Handbook of the Economics of Giving, Altruism, and Reciprocity,* vol. 2, edited by Serge Christophe Kolm and Jean Mercier Ythier. Amsterdam: North Holland, 1135 98. An ad vanced treatment of the determinants of remittances and their impacts on economic growth.

Ratha, Dilip. 2003. "Workers' Remittances: An Important and Stable Source of External Development Finance." *Global Development Finance: Striving for Stability in Development Finance*. Washington, DC: World Bank, 157 75. An often cited article on remittances in the context of financial flows for development.

Rosenzweig, Mark R., and Oded Stark. 1989. "Consump tion Smoothing, Migration, and Marriage: Evidence from Rural India." *Journal of Political Economy* 97 (4): 905 24. A classic article on the interplay of remittances and migration.

Woodruff, Christopher, and Rene Zenteno. 2007. "Mi gration Networks and Microenterprises in Mexico." *Journal of Development Economics* 82 (2): 509 28. An excellent treatment of migration, remittances, and small business development in Mexico.

World Bank. Monthly. World Bank Indicators online. Accessed May 2007. Available at http://web.worldbank .org/WBSITE/EXTERNAL/DATASTATISTICS.

SUSAN POZO

■ **rent seeking**

See political economy of trade policy

■ **research and development intensity**

See foreign direct investment (FDI)

■ **reserve currency**

A reserve currency is a foreign currency held by central banks or monetary authorities for the purpose of exchange intervention (to influence the exchange rate of the national currency) and the settlement of intergovernmental claims (debts). A reserve currency is also used in the pricing of goods, services, and assets entering international trade and finance. A reserve-currency status allows the nation issuing the currency to pay a somewhat lower price for imports (because it need not incur the small cost of exchanging its cur rency for another to pay for its imports), borrow at marginally lower rates, and earn seigniorage (the difference between the minimal cost of printing money and its face value).

The U.S. dollar is by far the largest vehicle cur rency in the world today. The euro (the currency of the European Monetary Union, or EMU) has in creasingly been used as a vehicle currency since its creation in 1999. To be a vehicle currency, a national currency must (1) be stable in value; (2) belong to a nation that occupies a central, or at least a very im portant, position in world production, trade, and finance; and (3) be fully convertible into other cur rencies and freely tradable in a viable foreign ex change market.

Table 1 shows that the share of national currencies in official holdings of foreign exchange reserves in the

form of U.S. dollars rose from 62.1 percent in 1996 to 71.0 percent in 1999, declined to 65.8 percent in 2004, then rose to 66.5 percent in 2005. The Japanese yen was 6.7 percent of the total in 1966 but declined to 3.6 percent in 2005; the British pound was 2.7 percent in 1996 and rose to 3.7 percent in 2005; the Swiss franc's share was very small (0.3 percent) both in 1996 and in 2005 (0.1 percent). The euro accounted for 17.9 percent of official holdings of foreign currency reserves in 1999 (the year of its

Table 1

Share of national currencies in official holdings of foreign exchange reserves at year end

	1996	1998	1999	2000	2001	2002	2003	2004	2005
U.S. dollar	62.1	69.4	71.0	71.0	71.4	67.0	65.9	65.8	66.5
Japanese yen	6.7	6.2	6.4	6.1	5.1	4.4	3.9	3.8	3.6
Pound sterling	2.7	2.7	2.9	2.8	2.7	2.8	2.8	3.4	3.7
Swiss franc	0.3	0.3	0.2	0.3	0.3	0.4	0.2	0.2	0.1
Euro			17.9	18.4	19.3	23.9	25.3	25.0	24.4
Deutsche mark	14.7	13.8							
French franc	1.8	1.6							
Nether. guilder	0.2	0.3							
ECUs	7.1	1.2							
Other currencies	4.4	4.5	1.6	1.4	1.2	1.5	1.9	1.8	1.7
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 2

Currency composition of official holdings of foreign exchange reserves at year end (in million $)

	1996	1998	1999	2000	2001	2002	2003	2004	2005
U.S. dollar	760.1	888.7	977.4	1,077.6	1,117.7	1,202.2	1,465.9	1,737.2	1,869.9
Japanese yen	82.3	80.0	87.8	93.3	79.4	78.4	87.8	101.5	100.7
Pound sterling	32.9	34.1	39.8	41.8	42.4	50.5	61.6	89.2	105.3
Swiss franc	3.7	4.2	3.2	4.1	4.4	7.3	4.9	4.3	4.0
Euro			246.9	279.5	301.9	428.1	562.4	660.0	685.4
Deutsche mark	179.9	176.9							
French franc	22.5	20.8							
Nether. guilder	2.9	3.5							
ECUs	86.6	15.3							
Total of Above	1,170.9	1,223.5	1,355.1	1,496.3	1,545.8	1,766.5	2,182.6	2,592.2	2,765.3
Other currencies	53.2	57.4	21.6	22.2	19.5	27.0	42.8	48.0	44.9
Overall total[a]	1,566.2	1,643.8	1,783.7	1,942.6	2,053.0	2,409.1	3,029.2	3,749.4	3,171.9

Source: IMF, *Annual Report* (2006).

[a] Includes foreign exchange reserves whose currency composition could not be determined.

creation) and rose to 25.3 percent in 2003; it was 24.4 percent in 2005. This was less than the EMU's share of world output and trade but much larger than the sum of the share of the deutsche mark, French franc, Netherlands guilder, and ECU foreign currency reserves that it replaced. The ECU (the European Currency Unit) was introduced in March 1979 by the European Monetary System (EMS), a forerunner of the EMU. On December 31, 1998, ECUs were unwound into gold and dollars. Table 2 gives the dollar amount of the various foreign exchange reserves from 1996 to 2005.

Before World War I, gold was the only international reserve held by nations and the British pound was the only important international currency. Under the Bretton Woods system, which operated from 1947 until 1971, the value of the U.S. dollar was fixed in terms of gold and the international value of other currencies were then fixed in terms of dollars. The fact that U.S. dollars held by foreign central banks were redeemable in gold made the dollar "as good as gold." In time, a few other currencies joined the dollar as international currency reserves. Thus the Bretton Woods system was a gold-exchange standard, with both dollars and international currency reserves held as official reserves by nations' central banks and monetary authorities.

After World War II, the dollar replaced the British pound as the most important international reserve currency, to reflect the much greater importance of the United States than the United Kingdom in world output, trade, and finance, and also because the dollar better satisfied the other conditions for serving as an international reserve currency. The British pound, however, remained a significant international reserve currency after World War II because of tradition (recognition) and inertia, but its importance declined sharply after the mid-1970s when the pricing of petroleum and other primary commodities switched from British pounds to U.S. dollars.

In August 1971, the dollar was devalued (the dollar price of gold increased) and its convertibility into gold was abolished, thus ushering the world into a pure paper (essentially a de facto dollar) standard.

Early in the 21st century, foreign exchange reserves constitute the bulk of international reserve assets, with the dollar by far the largest and the euro second largest. Although gold remains an important international reserve asset, its relative importance has greatly diminished. Other official reserve assets are nations' net IMF positions (the amount that a nation could borrow without questions asked by the International Monetary Fund, or IMF) and special drawing rights, or SDRs.

In the present managed exchange rate system, the importance of international reserves ought to be less than under the previous fixed exchange rate system because a nation facing a balance of payments deficit can allow its currency to depreciate, thus correcting its deficit without the need to continuously finance it with international reserves unless more drastic forms of adjustment become necessary, such as a devaluation of the currency or import restrictions.

See also balance of payments; Bretton Woods system; convertibility; dollar standard; dominant currency; euro; exchange rate regimes; foreign exchange intervention; gold standard, international; International Monetary Fund (IMF); international reserves; money supply; reserve currency; seigniorage; special drawing rights; sterilization; vehicle currency

FURTHER READING

Bank for International Settlements. *Annual Report*. Basel: BIS, 2006. Report on the world's financial and banking situation, including the use of reserve currencies.

International Monetary Fund. *Annual Report*. Washington, DC: IMF, 2006. An analysis of the world's economic and financial situation, it includes a discussion of the relative use of various vehicle currencies.

Mundell, Robert A. 2000. "A Reconsideration of the Twentieth Century." *American Economic Review* (June): 327 40. A review and interpretation of the operation of the international monetary system during the past century by a Nobel Prize recipient in economics.

Salvatore, Dominick. 2007. *International Economics*. 9th ed. Hoboken, NJ: Wiley, chapter 21. An overview of the functioning of the international monetary system and the use of reserve currencies.

Taiwan, Government of. http://investintaiwan.nat.gov.tw/
en/env/stats/foreign exchange.html. Data on the for
eign exchange reserves held by each country.

Widdershins. "Reserve Currency." Available at http://
mg.blogs.com/reserve. A discussion of the role of a re
serve currency in international finance.

DOMINICK SALVATORE

■ revealed comparative advantage

Revealed comparative advantage (RCA) is an empirical measure of the extent to which a given country specializes in the export of a particular product or range of products, compared with a reference set of countries. It is usually computed from trade data.

International trade theory starts from the premise that production specialization across different countries, followed by exports of the goods specialized in and imports of goods specialized in by other countries, leads to overall welfare gains for all economies. The pattern of specialization is called "comparative advantage," which is the result of relative advantage in access to materials, energy, or technological skills that are crucial for the efficient production of the exported good. Other countries would establish relative advantages in other goods that they export. The patterns of exports and imports would then derive from this relative or comparative advantage in production.

Trade theory begins by predicting patterns of trade from the existence of relative or comparative differences in natural resource endowments or capability acquired from production experience or technological change. The RCA approach starts from the other end of the production-trade chain, that is, the empirically observed patterns of export and import specialization revealed by the trade data. The explanation of the pattern of specialization is treated as a secondary matter to be inferred from the observed data on trade. Thus if the proportion of wine in France's exports is higher than the ratio of wine to all exports in the Organisation for Economic Co-operation and

Development (OECD) group as a whole, France is said to have revealed comparative advantage in the export of wine relative to the OECD reference set. Note that this does not imply that France would have RCA with respect to some other set of countries.

Origin and Mathematical Formulation The index most commonly used to measure RCA is the Balassa index (BI), which was developed and popularized by Bela Balassa (1965, 1989), though the concept of a specialization index was first proposed by Liesner (1958). If we wish to examine whether country i has manifest advantage in the export of product line j (at some level of product aggregation) over a set of reference countries represented by the symbol w, we construct the following simple Balassa index:

$$BI_{ij} = \textit{share of j in country i's exports/} \atop \textit{share of j in set w's exports} \quad (1)$$

If BI_{ij} is greater than 1.0, then country i would have RCA in product line j with respect to reference set w, since a proportionately greater share of its exports consist of j. If BI_{ij} is less than 1.0, the country is said to have revealed comparative disadvantage in good j. If we wish to be precise, we would have to specify the set w in the index as well, say as BI_{ijw}, since the index depends on the particular set of countries chosen. Exports are measured as total export values to the rest of the world expressed in a particular year's currency units. Clearly this index is very simple to compute from the actual trade data for any level of aggregation of export lines and is intuitively meaningful.

The index can also be extended from a single country i to a set k within a larger reference set w. For example, if we wish to compute the RCA of advanced industrial countries (AIC) in the export of capital goods, we would need to first aggregate all capital goods exported by all AICs and compute the share of these in total exports. This ratio would then have to be divided by the share of aggregated capital goods in total world exports. The index can be similarly modified to track the relative strength of any given trading region with respect to any particular characteristic of

exports, provided that it can be quantified, measured, and aggregated across a range of product lines.

In formal terms the Balassa, or revealed comparative advantage, index is often defined as follows for exporting region i, export line j and reference region w, where X_{mn} stands for export value of good n from region m and t stands for all exports from the particular region. The reference set w is often taken as the whole world:

$$RCA_{ij} = (X_{ij}/X_{it})/(X_{wj}/X_{wt})$$
$$= (X_{ij}/X_{wj})/(X_{it}/X_{wt}) \qquad (2)$$

Practical Issues Since its initial formulation, RCA indexes have been widely used to track specialization with regard to participation in the international economy. Starting from specialization in particular goods or ranges of goods, RCAs have been calculated for goods intensive in capital or labor, intermediate products, energy, a wide range of agrobased exports, and exports that use particular technologies. It is often the "first cut" analytical tool in export competitiveness studies carried out by academics and international consultants.

RCAs are even worked out for the generation of patentable innovations (Laursen 1998). The RCA concept has turned out to be a very flexible tool of analysis and the original formula has been modified in a wide variety of ways, a few of which are outlined in the next section. It is also clear that the formulation can be adapted easily for import specialization as well and there are indexes that use both export and import data.

RCAs must be used with some care. They are useful to identify sectors that are strong and have the potential to be built up further. But the index cannot be used to infer competitive strength vis-à-vis another country in the same reference set. A large country with a more diverse range of exports will have a smaller RCA than a small nation that concentrates on a much smaller set of exports but has comparable competitive strength. Nor do relative values of different products for the same country say much about the relative importance of these products for the economy. For example, tobacco has the highest Balassa index for the United States over a long time

span, but this is hardly its most important export product (Marrewijk 2002).

RCA measures are also affected by trade barriers or subsidies since these could distort exports or raise production costs by raising the price of imported inputs. Reexports and a high degree of intraindustry trade also pose special problems, which can be handled by the use of alternative measures (Laursen 1998; Utkulu and Seyman 2004).

Extensions One problem with the basic Balassa index is that it is not symmetrically distributed around the neutral value 1.0, ranging from 0 to 1 for comparative disadvantage and indefinitely upward from 1.0 for comparative advantage products. This problem is easily corrected by taking natural logarithms of the ratios with the index defined as follows:

$$RCA_{ij} = \ln(X_{ij}/X_{it}) - \ln(X_{wj}/X_{wt}) \qquad (3)$$

The revised index is now symmetric around 0. This form is particularly useful for econometric studies. An alternative transformation, which yields RCA values between 1 and +1 around 0, is given in Laursen (1998), but it is probably less useful than the log form for econometric work.

Vollrath (1991) proposed the following alternative index of trade specialization, which brings imports into the calculation:

$$RCA_{v1} = RXA - RMA \qquad (4)$$

Where $RXA = (X_{ij}/X_{it})/(X_{wj}/X_{wt})$, the original export specialization index, and $RMA = (M_{ij}/M_{it})/(M_{wj}/M_{wt})$, which is a similar import specialization index.

An alternative formulation of the Vollrath index takes the following form:

$$RCA_{v2} = \ln RXA - \ln RMA \qquad (5)$$

The Michaely index MI (see Laursen 1998) for country i and product j is defined as the difference between the share of sector j in total exports minus the share of sector j in total imports. There is no reference set of countries, so the index ranges from 1 to +1 and is 0 if import shares are perfectly balanced with export shares:

$$MI_{ij} = X_{ij}/X_{it} - M_{ij}/M_{it} \qquad (6)$$

A particularly useful measure of normalized revealed comparative advantage (NRCA), proposed by Yu et al. (2008), is defined in equation 7, where w

refers to the entire world and t to total exports. This index can be interpreted as the deviation of the normalized value of country i exports of product line j with respect to total world exports (the first term) from its expected comparative advantage neutral value (the second term).

$$NRCA_{ij} = X_{ij}/X_{wt} - X_{wj} \cdot X_{it}/X_{wt}^2 \qquad (7)$$

The equivalent form of equation 8 shows an alternative deviation kernel (term one) normalized by the relative weight of country i exports in world exports (term two).

$$NRCA_{ij} = [X_{ij}/X_{it} - X_{wj}/X_{wt}]\{X_{it}/X_{wt}\} \qquad (8)$$

This index has some useful properties: it is symmetrical about zero and displays additive consistency across product lines and country groupings. It just needs to be scaled up by an arbitrary factor to facilitate numerical comparisons, since both terms in equation 7 yield very small values; but this is no problem since RCA is a relative measure in any case. Furthermore the sum of NRCA across all commodities for any one country is zero as is the sum of NRCA across all countries for any one commodity.

Overall the RCA concept has proved to be an enduring and easily computable measure that can be used to track patterns of trade and production specialization. It has been used to study the performance of individual countries or groups of countries with respect to a larger set or even to comprehensively compare countries with one another, as in Batra and Khan (2005), which compares the evolution of India and China. RCA is therefore a very useful and powerful tool of trade analysis, used by academics and economic consultants to study evolving patterns in the world economy.

See also comparative advantage; intraindustry trade

FURTHER READING

Balassa, Bela. 1965. "Trade Liberalization and 'Revealed' Comparative Advantage." *Manchester School of Economic and Social Studies* 33 (2): 92–123. The first published use of the Balassa index to measure RCA.

———. 1989. "'Revealed' Comparative Advantage Revisited." In *Comparative Advantage, Trade Policy, and Economic Development*, edited by Bela Balassa. New York: New York University Press, 63–79. In this paper, Balassa further develops RCA measures in trade analysis.

Batra, Amita, and Zeba Khan. 2005. "Revealed Comparative Advantage: An Analysis for India and China." Working Paper No. 168. New Delhi: Indian Council for Research on International Economic Relations. Downloadable from http://www.icrier.org. Illustrates the value of the RCA concept to track evolving patterns of export and production specialization in India and China.

Laursen, Keld. 1998. "Revealed Comparative Advantage and the Alternatives as Measures of International Specialization." DRUID Working Paper 98–30. Copenhagen. An exposition of various RCA measures that are compared for many countries.

Liesner, H. H. 1958. "The European Common Market and British Industry." *Economic Journal* 68 (270): 302–16. The first use of the relative share concept to measure specialization in trade.

Marrewijk, Charles Van. 2002. *International Trade and the World Economy*. Oxford: Oxford University Press, 36–40. Presents a simple introduction to the Balassa index along with practical applications.

Utkulu, Utku, and Dilek Seyman. 2004. "Revealed Comparative Advantage and Competitiveness: Evidence for Turkey vis à vis the EU–15." European Trade Study Group, 6th Annual Conference (ETSG 2004). Nottingham, UK. Employs alternative RCA indexes to examine Turkey's exports to the European Union and discusses the merits of various indexes.

Vollrath, T. L. 1991. "A Theoretical Evaluation of Alternative Trade Intensity Measures of Revealed Comparative Advantage." *Weltwirtschaftliches Archiv* 127 (2): 265–80. A theoretical assessment of the various RCA measures in use.

Yu R., J. N. Cai, and P. S. Leung. 2008. "The Normalized Revealed Comparative Advantage Index." *Annals of Regional Science*. Available online at http://www.springerlink.com/content/qx7j588366801n05/fulltext.pdf.

G. CHRIS RODRIGO

■ Ricardian model

The Ricardian model is the simplest and most basic general equilibrium model of international trade that

we have. It is usually featured in an early chapter of any textbook on international economics. Historically, it is the earliest model of trade to have appeared in the writings of classical economists, at least among models that are still considered useful today.

It is indeed still useful. In spite of being superseded over the years by models with much more complexity (more factors of production, increasing returns to scale, imperfect competition), the Ricardian model often provides the platform for the introduction of today's new ideas. Dornbusch, Fischer, and Samuelson (1977) examined a continuum of goods first in a Ricardian model. Eaton and Kortum (2002) incorporated an ingenious and elegant treatment of geography into a Ricardian model. Melitz (2003) started a small revolution in trade theory by modeling heterogeneous firms within what was essentially a Ricardian model.

The Ricardian model itself, as a new idea, came many years after Ricardo. According to Ruffin (2002), in 1816 David Ricardo introduced only a portion of the model that now bears his name, focusing primarily on the amounts of labor used to produce traded goods and, from that, the concept of comparative advantage. The first appearance of the Ricardian model, according to Ruffin again, was in Mill (1844).

The Simple Ricardian Model The simple Ricardian model depicts a world of two countries, A and B, each using a single factor of production, labor L, to produce two goods, X and Y. Technologies display constant returns to scale, meaning that a fixed amount of labor, a_g^c is needed to produce a unit of output of each good, $g = X,Y$, in each country, $c = A,B$, regardless of how much is produced in total. All markets are perfectly competitive, so that goods are priced at cost in countries that produce them, $p_g^c = w^c a_g^c$, where w^c is the competitive wage in country c. Labor is available in fixed supply in each country, L^c; it is immobile between countries but perfectly mobile within each. The Ricardian model typically leaves demands for goods much less fully specified than supplies, though a modern formulation may specify for each country a utility function, $U^c = U^c(C_x^c, C_y^c)$, which the representative con-

sumer maximizes subject to a budget constraint. Utility functions may, or may not, be assumed in addition to be identical across countries, or to possess other regularity properties, although most properties of the model's solution do not require any of these assumptions.

The most basic use of the model compares the equilibriums in autarky (that is, complete self-sufficiency without trade) with those of free and frictionless trade. In autarky, since both goods must be produced in each country, prices are given immediately by the costs stated above, and further analysis is needed only if one wants to know quantities produced and consumed. If so, the linear technology implies a linear production possibility frontier (PPF) that also serves as the budget line for consumers in autarky. The autarky equilibrium is as shown in figure 1, where $\tilde{~}$ indicates autarky and Q represents production.

Comparison of the two countries in autarky depends primarily on their relative costs of producing the two goods, which in this model defines their comparative advantage. For concreteness, assume that country A has comparative advantage in good X: $a_X^A/a_Y^A < a_X^B/a_Y^B$, so that $\tilde{p}_X^A/\tilde{p}_Y^A < \tilde{p}_X^B/\tilde{p}_Y^B$. Without further assumptions about preferences, little more can be said about autarky, but with addi-
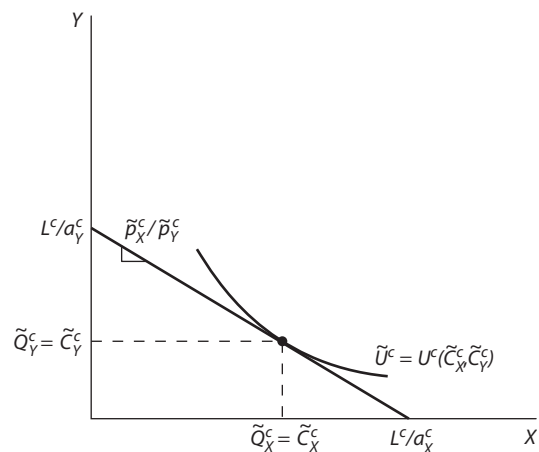
Figure 1
Ricardian model equilibrium in autarky

tional assumptions about preferences, one can infer that $\tilde{Q}_X^A/\tilde{Q}_Y^A > \tilde{Q}_X^B/\tilde{Q}_Y^B$.

With free and frictionless trade, prices must be the same in both countries. Two kinds of equilibrium are possible, depending on the supplies and demands for goods in the two countries. One kind of equilibrium has world relative prices, denoted here by ˇ, strictly between the relative prices of the two countries in autarky: $\tilde{p}_X^A/\tilde{p}_Y^A < \check{p}_X/\check{p}_Y < \tilde{p}_X^B/\tilde{p}_Y^B$. In that case, each country must specialize in producing only the good for which its relative cost is lower than the world relative price, thus the good in which it has comparative advantage. Each must necessarily export that good.

With such complete specialization, outputs of the goods are determined by labor endowments and productivities, so equality of world supply and demand must be achieved from the demand side. That is, world prices are determined such that the two countries' demands sum to the quantity produced in one of them. These demands derive from the expanded budget constraints of each country's consumers, reflecting the value at world prices of the single good that the country produces. Consumers can now, unless they wish to consume only that single good, consume more of both goods than they did in autarky. Whether they choose to do so depends on the extent to which they substitute toward the cheaper good now imported from abroad, but in any case they reach a higher indifference curve and are better off. All of this is shown in figure 2. For this to be an equilibrium, the quantity of each good exported by one country must equal the quantity imported by the other, so the heavy arrows showing net trade in each panel of the figure must be equal and opposite.

Such an equilibrium with specialization will arise only if the two countries' capacities to produce their respective comparative-advantage goods correspond sufficiently closely to world demands for the goods. If this is not the case   if one country's labor endowment is too low and/or its labor requirement for producing its comparative-advantage good is too high for it to satisfy world demand   then while that country will specialize, the other country (call it the larger one, although that is not strictly necessary) will not. Instead of world relative prices settling between the two autarky levels as described earlier, prices will exactly equal the autarky prices of the larger country, and that country will produce both goods. At those prices, producers in the larger country will be indifferent among all output combinations on the PPF,
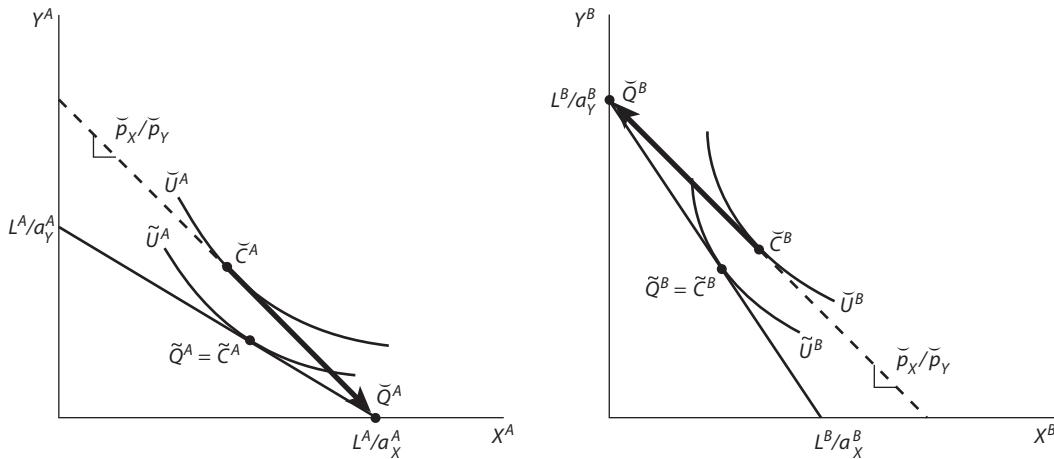
Figure 2

Free trade equilibrium with complete specialization

and output in the large country will be determined instead by the need to fill whatever demand is not satisfied by the smaller country.

Such an equilibrium is shown in figure 3, where in comparison to figure 2 country B's labor endowment has been made smaller and both countries' preference for good Y has been increased. As a result, country B is too small to meet world demand for good Y, even at country A's autarky prices. Therefore the free trade equilibrium has country A consuming where it did in autarky, while its production, $\breve{Q}^A$, moves down along its PPF so that, again, its trade vector can be equal and opposite to that of country B. Note that, in this trading equilibrium, the larger country neither gains nor loses from trade.

The following are some of the implications of this simple model, some of which have been illustrated above, while others can be derived rather simply:

Effects of trade:

- Each country exports the good in which it has comparative advantage, as defined by having a lower relative autarky price than the other country.
- Trade causes each country to expand its production of the good it exports, with labor being reallocated to it from the import-competing industry.

- Trade causes the relative price of a country's export good to rise, except in the case of a "large" country, defined here as one whose trading partner is too small to meet its demand for imports.
- Consumption and welfare are unchanged by trade in a large country; in any country that is not large, consumers buy more of one or both goods and welfare increases.
- Because all income accrues to labor, which earns the same wage in both industries due to mobility, conclusions about welfare or utility apply equally well to the real wage.

Effects of changes in trading equilibriums (assuming a bit more about preferences):

- An increase in the labor endowment of a country, holding other labor, technology, and tastes constant, hurts the growing country and benefits the other.
- A fall in the labor required by a country to produce its export good, holding other technology, endowments, and tastes constant, benefits the other country but may either benefit or harm ("immiserize") the growing country.
- A rise in the labor required by a country to produce its import good has no effect if it does not produce that good; if it does pro-
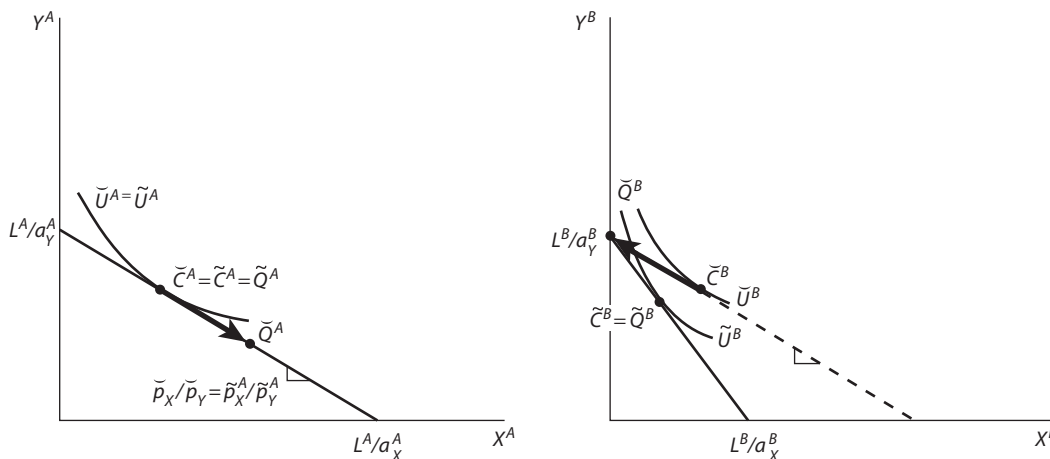
Figure 3

Free trade equilibrium with country A incompletely specialized

duce it (like country A in figure 3), the world price of that good rises, that country is harmed, and the other country gains.

- A change in preferences, in either country, in favor of one of the goods has no effect on prices or production if one of the countries is incompletely specialized. If both are specialized, however, then the relative price of that good rises, improving the terms of trade of the country that exports it.

Extensions of the Simple Ricardian Model Before considering several extensions of the simple model described here, it is reasonable to ask what extensions would *not* be acceptable, in that they would lead to a model that would no longer be "Ricardian," as trade economists understand the term. Ricardo himself might disagree, were he alive, but the essential features of a Ricardian model seem to be two: that production uses only homogeneous labor as a primary input and that comparative advantage arises from differences across goods and countries in the technology for producing goods from that labor. Both of these requirements distinguish a Ricardian model from the other principal model of trade theory, the Heckscher-Ohlin or factor-proportions model. With primary factors other than labor (including different kinds of labor based on skill and/or industry of location), a model takes on features that differ in essential ways from the Ricardian model. On the other hand, with only homogeneous labor as a factor of production, if technologies do not differ across countries then there is no scope for comparative advantage based trade.

More Goods and/or Countries Therefore, keeping the number of factors at one, the most obvious things to extend in the simple model are to add to the numbers of goods and/or countries. This is relatively easily done, as long as one does not try to do both.

With two countries and many goods, the goods can be ranked in a chain of comparative advantage based on the ratios of their unit labor requirements in the two countries. That is, if one numbers N goods such that $a_1^A/a_1^B < a_2^A/a_2^B < \cdots < a_N^A/a_N^B$, then country A has comparative advantage in the low end of this ranking while country B has it in the high end.

Further one can show that under free trade, each country will specialize in and export goods in its respective end of the chain, with at most one good (and perhaps no good) being produced in common by both countries. The division between A's exports and B's exports depends on country sizes, technologies, and tastes, much as in the choice between figures 2 and 3. For example, the larger the labor endowment and/or efficiency of country A compared with B, the further up the chain will A produce and export.

Similarly, with two goods and many countries, the countries can be ranked in a chain of comparative advantage based on the ratios of their unit labor requirements for producing the two goods. That is, if one numbers M countries such that $a_X^1/a_Y^1 < a_X^2/a_Y^2 < \cdots < a_X^M/a_Y^M$, those countries in the low end of this ranking will specialize in and export good X to those in the high end, which export good Y. Again there will be at most one country (and perhaps no country) that produces both goods. And the division between X-exporters and Y-exporters depends on country sizes, technologies, and tastes.

Unfortunately, extending to more than two of both goods and countries is not so simple or intuitive. Jones (1961) seems to have done about as well as one can, showing that an efficient assignment of countries to goods will minimize the product of their unit labor requirements. This certainly suggests the importance of comparative advantage, in the form of low relative unit labor requirements, which is perhaps all that one should hope for from a many-good, many-country Ricardian model (though see Eaton and Kortum's solution to this problem, which we discuss later).

A Continuum of Goods A less obvious, but much more useful, extension of the Ricardian model was provided by Dornbusch, Fischer, and Samuelson (1977) hereinafter DFS who took the number of goods to infinity, in the form of a continuum. Indexing goods by the continuous variable j on the interval [0,1], they specified technologies for each of two countries as $a^c(j)$ representing the amount of labor required in country c to produce one unit of good j. The ratio of these in the two countries of the model, ordered monotonically in a function $A(j)$,

then plays the same role as the chain of relative labor requirements mentioned earlier for the many goods case. But with a continuum of goods, the good at the dividing line between a country's exports and its imports is of negligible importance for labor markets, since it employs a negligible amount of labor, and this removes the need to consider whether a good is produced in both countries. Such a good now always exists, as the dividing line between one country's exports and its imports, but it is of negligible importance for employment.

This simplicity is helpful in itself, but the more important advantage of the continuum model is that it facilitates the analysis of the range of goods that a country will export and import, something that the two-good model could not usefully address. One finds, for example, that an expansion of the labor endowment of one country relative to the other will cause it to expand its exports, not just by exporting more of what it already exported (though that happens too), but by exporting goods that it previously imported.

The model in its simplest form is depicted in figure 4, which is taken directly from DFS's figure 1. The downward sloping function $A(j)$ is the ratio of the two countries' unit labor requirements, ordered so that country A's comparative advantage declines with rising j. Letting $\omega = w^A/w^B$, for any given value of this relative wage, free and frictionless trade will lead to country A producing and exporting all goods with $A > \omega$ and importing goods with $A < \omega$. To determine the equilibrium value of ω one needs assumptions about demand, which are reflected in the upward sloping curve $B(j; L^B/L^A)$. Assuming that preferences are identical and that constant shares of expenditure are spent on each good, this curve measures the relative wage at which demands for each country's range of goods produced would equal their supplies (or, equivalently, the relative wage at which values of a country's exports and imports will be equal). This requires simply that the ratio of expenditures on the two sets of goods equals the ratio of the incomes of those who produce them. As the definition of this market-clearing relative wage shown in figure 4 indicates, it depends positively on $\vartheta(j)$, the

fraction of income spent on the goods produced by country A, which in turn rises with the fraction of goods that A produces. It also depends positively on the relative size (labor force) of country B, since the larger that is, for a given division of goods between the two countries, the higher must be the relative wage in A to keep the expenditure ratio constant.

Figure 4 immediately yields the result mentioned earlier, that as a country's labor force rises relative to the other country (shifting the B curve up or down), its share of goods produced increases as well, while its relative wage falls. Likewise, if a country becomes more productive in producing all goods (its $a^i(j)$ shifts down, shifting the A curve up or down), it also produces more goods but its relative wage increases. Other exercises are possible with the simple model, and DFS extend the model in a variety of directions to illuminate many issues that could not be readily addressed in models with a finite number of goods.

Most notably, they incorporate transportation costs, giving rise to a third endogenous range of goods, in addition to those exported by countries A and B: nontraded goods, whose costs differ too little between countries to overcome the barrier of transportation costs. This is particularly useful, since it
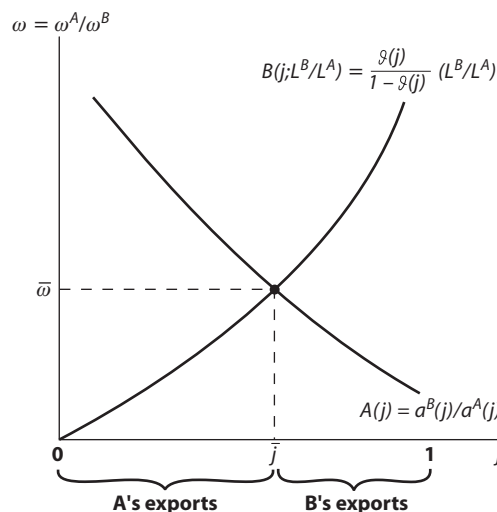
Figure 4

Ricardian model with a continuum of goods

implies that as a country's relative productive capacity rises, some of the goods it previously imported become nontraded, while it begins to export some of the goods that were previously nontraded.

Multiple Countries with Random Technologies A limitation of the DFS model is that it applies to a world of only two countries, and because of its reliance on ratios of values in those countries it is not readily extended to more, although some (e.g., Wilson 1980) have had some success. A breakthrough was provided by Eaton and Kortum (2002), however, who extended the DFS model to an arbitrary number of countries by assuming that, in effect, the labor productivities of each good and country are determined randomly. Specifically, they let labor productivity, $z^i(j) = 1/a^i(j)$, be determined by a random draw from a probability distribution, such that each country has some probability, regardless of its overall technical ability and its wage, of having a lower cost than any other country. This probability, then, translates into the fraction of the continuum of goods that the country is able to produce and export under free and frictionless trade. More important, by including transportation costs for each pair of countries, each country has a fraction of goods that it will be able to produce even without necessarily having the lowest costs, since they only need to cost less than goods from other countries inclusive of transportation cost. Furthermore if transportation costs are low enough that a country imports anything, then it will also export some fraction of goods, since if necessary the wage will fall until some fraction of goods can be exported to one or more countries for delivered prices below those countries' domestic prices. This formulation therefore extends the Ricardian model not only to multiple countries but to a context that can account for bilateral trade.

The Eaton-Kortum model generates equations for prices and trade shares that provide the basis for empirical estimation as well as being susceptible to solution and comparative static analysis by numerical methods. The model provides an elegant and parsimonious theoretical justification for the gravity model of bilateral trade flows while illustrating the interaction between the forces of comparative advantage that give rise to trade and the geographical resistance to those forces in the form of transportation and other costs of trade that limit trade and direct it over particular geographical routes.

The Role of the Ricardian Model in Understanding the World Economy The Ricardian model was introduced long ago to explain one of the most basic concepts of economics, comparative advantage, but it continues to be useful as a framework for understanding how countries interact in trade. Countries continue to differ in their abilities to produce goods and services, and the extent of these differences varies across products. This is not the only reason for international trade, but it is surely the most basic. Modern extensions of the Ricardian model are essential for understanding this trade and anticipating its effects.

See also absolute advantage; comparative advantage; gravity models; Heckscher-Ohlin model; New Trade Theory; specific-factors model

FURTHER READING

Collins, Susan M. 1985. "Technical Progress in a Three Country Ricardian Model with a Continuum of Goods." *Journal of International Economics* 19 (1 2): 171 79. Shows in a Ricardian model that technical progress in one country can hurt another whose exports are close substitutes for the first.

Deardorff, Alan V. 2001. "Fragmentation in Simple Trade Models." *North American Journal of Economics and Finance* 12 (2): 121 37. Examines the splitting of production processes across countries in a Ricardian model.

Dornbusch, Rudiger, Stanley Fischer, and Paul A. Samuelson. 1977. "Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods." *American Economic Review* 67 (5): 823 39. The seminal paper extending the Ricardian model to an infinite number of goods, with additional extensions to include transport costs, tariffs, money, and exchange rates.

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography, and Trade." *Econometrica* 70 (5): 1741 79. Uses randomized technologies to extend the Ricardian model to multiple countries and a continuum of goods;

rich with both theoretical implications and empirical applications.

Hicks, John R. 1953. "An Inaugural Lecture." *Oxford Economic Papers* 5 (2): 117 35. Includes a classic treat ment of the effects of technological progress in a Ri cardian model.

Jones, Ronald W. 1961. "Comparative Advantage and the Theory of Tariffs: A Multi Country, Multi Commodity Model." *Review of Economic Studies* 28 (3): 161 75. Extends the predictions of comparative advantage to a model with multiple (but finite numbers of) goods and countries.

Melitz, Marc J. 2003. "The Impact of Trade on Intra industry Reallocations and Aggregate Industry Pro ductivity." *Econometrica* 71 (6): 1695 25. Seminal paper modeling heterogeneous firms in a one factor trade model.

Mill, John Stuart. 1844. "Of the Laws of Interchange be tween Nations; and the Distribution of the Gains of Commerce among the Countries of the Commercial World." *Essays on Some Unsettled Questions of Political Economy*. London: John W. Parker. Classic treatment of international trade that includes the first full specifica tion of the Ricardian model.

Ruffin, Roy J. 2002. "David Ricardo's Discovery of Com parative Advantage." *History of Political Economy* 34 (4): 727 48. Documents the process by which David Ri cardo discovered the idea of comparative advantage, including his thinking that anticipated much of modern trade theory.

Stern, Robert M. 1962. "British and American Productivity and Comparative Costs in International Trade." *Oxford Economic Papers* 14 (3): 275 96. A classic empirical test of the Ricardian model.

Wilson, Charles A. 1980. "On the General Structure of Ricardian Models with a Continuum of Goods." *Econ ometrica* 48 (7): 1675 1702. Extends the Dornbusch Fischer Samuelson model to multiple countries.

ALAN V. DEARDORFF

■ **rules of origin**

Rules of origin (ROOs) specify the conditions under which a good is deemed to originate from a country and hence is eligible for preferential treatment. Such treatment could involve zero or preferential tariffs in a free (or preferential) trade area (FTA), or the right to export under a country-specific quota as under the old Multifiber Arrangement regime in textiles and clothing. Since there are zero (or preferential) tariffs on goods imported from member countries of an FTA, but tariffs on nonmembers differ across coun tries, the opportunities for arbitrage are obvious: imports into the area as a whole should take the path of least resistance, that is, enter through the country with the lowest tariffs on them and flow freely from there into other FTA member markets. As a result the lowest-tariff country would gain tariff revenue at the cost of other members. This suggests that countries would compete to set the lowest tariffs in order to attract such trade flows (see Richardson 1995). This seems to have actually happened in the United States after the colonies obtained independence from the British but before they fully integrated and set a common tariff (see McGillivray and Green 2001; Viner 1950). ROOs are seen as limiting such arbi trage possibilities.

Defining ROOs There are four common criteria for defining rules of origin.

1. Domestic content: content can be defined in terms of value added or in physical terms.
2. Change in tariff heading (CTH): a product must change its tariff heading (i.e., where it is classified in the tariff code) in a specified way within the FTA or customs union (CU) to obtain origin and hence be eligible for the preferential treatment reserved for the country.
3. Specified process: this outlines the processes that must be performed within the FTA or CU to obtain origin. Restrictiveness de pends entirely on the steps prescribed and the nature of the production technology. The difference between this and the CTH criterion is only that the latter is based on some commonly used descriptions such as the tariff code, whereas the specified process definition is defined in terms of production processes specific to each industry.

4. Substantial transformation: this is more loosely defined. In the United States the term *substantial transformation* has come to mean the determination of origin based on common law, reasoning from case to case. This process results in a commodity-specific ROO that falls into one of the first three categories.

Origin requirements can be made more restrictive by requiring more than one criterion to be met. Exceptions can also be used to make the origin requirement more restrictive. For example, under the North American Free Trade Agreement (NAFTA), transformation from any other chapter (two-digit classification level) of the harmonized system to tomato catsup, chapter 21, confers origin *except* transformation from tomato paste, which falls in chapter 20!

Effects of ROOs Although it is tempting to think of ROOs as necessary bureaucratic nuisances, they are often far more. First, ROOs raise the costs of exports when they force a firm to use a particular process or input to obtain origin. This cost increase can significantly erode the benefits supposedly offered to developing country exports through special trade preferences. For example, Mattoo et al. (2002) argue that the value of concessions by the United States under the African Growth and Opportunity Act are cut in half by stringent ROOs. In addition, ROOs are often quite expensive to document, as detailed records need to be kept and certification procedures may need to be followed. As a result, even if a product satisfies origin, an importer may prefer to pay the tariff rather than bother with the documentation needed. Herin (1986) shows that the cost of proving origin seems to have led more than a quarter of European Free Trade Area (EFTA) exports to pay the tariff despite the possibility of not having to do so. These costs are a pure waste of resources and work to reduce aggregate income and welfare.

Second, ROOs are, in themselves, a form of hidden protectionism: they provide an incentive for regional producers to buy intermediate goods from regional sources, even if their prices are higher than those of the identical import from outside the FTA, in order to make their product originate in the FTA and qualify for preferential treatment. This, in effect, protects FTA suppliers. As a result, trade patterns and investment flows needed to sustain them can be profoundly affected by a FTA. Consider a producer of shirts in Bangladesh. If he uses imported fabric the shirt would not meet the ROOs under the EBA (everything but arms) initiative and be eligible for zero tariffs in the European Union (EU). Only when cloth made from Bangladeshi yarn (the so-called yarn forward rule) is used is origin granted. Otherwise, the most-favored-nation tariff (of about 12 percent to 15 percent) is applied. As a result, the shirt maker is better off using more expensive Bangladeshi cloth, so long as the Bangladeshi cloth does not raise his costs by more than the amount he saves by avoiding the tariff. In this manner, the ROOs act as a hidden subsidy for domestic cloth and a tax on imported cloth; small wonder that Bangladeshi textile producers support such ROOs, while apparel exporters oppose them (Demidova, Kee, and Krishna 2006).

Third, as ROOs are negotiated industry by industry during the negotiations for the agreement, they are both hard to change ex post and allow enormous scope for well-organized industries to essentially insulate themselves from the effects of the FTA by devising suitable ROOs. The fact that ROOs are negotiated in such an industry-specific manner may be exactly why FTAs are so popular; however, they allow organized interest groups to evade the FTA by setting strict ROOs, thereby removing their opposition to the FTA. Grossman and Helpman (1995), for example, argue that being able to exclude certain sectors (what appropriately constructed ROOs will do) can make an FTA viable.

That ROOs may be protectionist does not prove that they are. Work by Estevadeordal (2000) suggests, however, that ROOs are being used to prevent trade deflection, as the sectors that have large differences in tariffs between the partners are the ones where ROOs are strongest. The work of Anson et al. (2005) on NAFTA also suggests that ROOs negate the effects of tariff reductions due to an FTA. They show that while the severity of ROOs reduced

Mexican exports, tariff preferences raised them, and the net effect was close to zero.

As ROOs are extremely opaque and arcane, they are easily overlooked, and given that there is much potential for abuse, the cost of overlooking them may be substantial.

See also customs unions; free trade area; regionalism

FURTHER READING

Anson, José, Olivier Cadot, Antoni Estevadeordal, Jaime de Melo, Akiko Suwa Eisenmann, and Bolormaa Tu murchudur. 2005. "Rules of Origin in North South Preferential Trading Arrangements with an Application to NAFTA." *Review of International Economics* 13 (3): 501 17. An early paper showing how ROOs limit the market access NAFTA seems to give.

Demidova, Svetlana, Hiau Looi Kee, and Kala Krishna. 2006. "Do Trade Policy Differences Induce Sorting? Theory and Evidence from Bangladeshi Apparel Ex porters." NBER Working Paper No.12725. Cam bridge, MA: National Bureau of Economic Research. An early paper that shows how in a heterogeneous firms setting, differences in effective preferences given to less developed countries by developed ones results in firms sorting themselves across markets.

Estevadeordal, Antoni. 2000. "Negotiating Preferential Market Access: The Case of the North American Free Trade Agreement." *Journal of World Trade* 34: 141 200. An early empirical approach that constructs an index of the restrictiveness of ROOs and relates it to preference margins and other variables.

Grossman, Gene, and Elhanan Helpman. 1995. "The Politics of Free Trade Agreements." *American Economic Review* 85 (4): 667 90. A classic paper that develops the use of the menu auction framework to understand the political economy behind the formation of FTAs.

Herin, Jan. 1986. "Rules of Origin and Differences between Tariff Levels in EFTA and in the EC." EFTA Occasional Paper No. 13. Geneva: European Free Trade Associa tion. Provides evidence in support of significant docu mentation costs in FTAs.

Krishna, Kala, and Anne Krueger. 1995. "Implementing Free Trade Areas: Rules of Origin and Hidden Protec tion." In *New Directions in Trade Theory*, edited by Alan Deardorff, James Levinsohn, and Robert Stern. Ann Arbor: University of Michigan Press, 149 87. Shows how ROOs affect trade and investment flows in a simple analytical model.

Krueger, Anne O. 1999. "Free Trade Agreements as Pro tectionist Devices: Rules of Origin." In *Trade, Theory, and Econometrics: Essays in Honor of John Chipman. Routledge Studies in the Modern World Economy*, edited by James R. Melvin, James C. Moore, and Raymond Riezman. London: Routledge, 91 102. An early paper showing how ROOs act as hidden protection.

Mattoo, Aaditya, Devesh Roy, and Arvind Subramaniam. 2002. "The Africa Growth and Opportunity Act and Its Rules of Origin: Generosity Undermined?" *World Economy* 26 (6): 829 51. A good example showing how preferences may not be doing much for developing countries.

McGillivray, Fiona, and Matt Green. 2001. "Trading in a Free Trade Area with No Rules of Origin: The U.S. Under the Articles of Confederation." Mimeo. New Haven, CT: Yale University. This provides a neat his torical example of an FTA without ROOs.

Richardson, Martin. 1995. "Tariff Revenue Competition in a Free Trade Area." *European Economic Review* 39: 1429 37. A simple model showing how tariffs might be competed down to get revenue in FTAs.

Viner, J. 1950. "The Customs Union Issue." New York: Carnegie Endowment for International Peace. A classic work on customs unions.

KALA KRISHNA

■ Rybczynski theorem

See Heckscher-Ohlin model

S

▪ safeguards

The term *safeguards* refers to trade intervention by national governments to address the injurious effects of import surges on import-competing industries. Trade intervention for this purpose has been subject to discipline under the General Agreement on Tariffs and Trade (GATT) since its inception in 1947 pursuant to GATT Article XIX. During the Uruguay Round of GATT negotiations, which culminated in the creation of the World Trade Organization (WTO) in 1994, discipline over safeguard measures was expanded and deepened through a new WTO Agreement on Safeguards. In addition, Article X of the General Agreement on Trade in Services authorizes negotiations over the possibility of safeguards in service sectors, but those negotiations have not yet resulted in any agreement. Consequently, safeguards are confined to goods sectors.

The Legal Foundation for Safeguards and Its Evolution in the WTO GATT Article XIX(1) provides:

> If, as a result of unforeseen developments and of the effect of the obligations incurred by a contracting party under this Agreement, including tariff concessions, any product is being imported into the territory of that contracting party in such increased quantities and under such conditions as to cause or threaten serious injury to domestic producers in that territory of like or directly competitive products, the contracting party shall be free, in respect of such product, and to the extent and for such time as may be necessary to

> prevent or remedy such injury, to suspend the obligation in whole or in part or to withdraw or modify the concession.

Article XIX thereby establishes four prerequisites for the use of safeguards. First, imports of the good(s) in question must have increased. Second, the increase must be a result of "unforeseen developments" and the "effect of the obligations incurred" under GATT. Third, the import-competing domestic industry must be suffering "serious injury" or a threat of such injury. Finally, a causal linkage must exist between the import surge (or the underlying developments that precipitated it) and the present or threatened injury.

Paragraphs two and three of Article XIX further provide that a party invoking its right to suspend or modify concessions must negotiate with adversely affected parties over the possibility of compensatory trade concessions. If these negotiations are unsuccessful, safeguards may be imposed nonetheless, but adversely affected trading partners then have the right to suspend "substantially equivalent concessions" in response. Thus the original structure of Article XIX permitted the use of safeguards but only at a price the GATT bargain was to be rebalanced either through an offer of trade compensation (the preferred outcome) or through a retaliatory withdrawal of substantially equivalent concessions.

Over the years, formal reliance on Article XIX within the GATT system diminished. Safeguard measures were largely replaced by what came to be known as "gray area" measures, such as voluntary restraint agreements and orderly marketing

agreements. These agreements were typically negotiated on a bilateral basis with major suppliers.

The proliferation of gray area measures was viewed as a serious problem by many observers. Gray area measures were not time-limited in the manner contemplated by Article XIX(1) and were employed regardless of whether the prerequisites for safeguards under Article XIX were met. In addition, gray area measures were often discriminatory in their impact, as less significant or less competitive suppliers were often exempted. A constituency thus developed for reform, which culminated in the Agreement on Safeguards in the Uruguay Round.

A key achievement of this agreement was a prohibition on gray area measures in Article XIII, coupled with a requirement for the phase out of existing gray area measures. In addition, to make reliance on formal safeguards more attractive, Article VIII(3) eliminated the rebalancing requirement (compensation or retaliation) for the first three years of any safeguard measure imposed in compliance with the agreement following an absolute increase in imports (as distinguished from a mere increase in their market share).

Other significant features of the agreement include a general prohibition on discriminatory safeguard measures, along with numerous transparency and procedural requirements applicable to any decision by national authorities to employ safeguards. The agreement limits the duration of safeguards to a maximum of eight years and provides that members proposing to employ them for more than four years must demonstrate anew that the prerequisites for safeguard measures are satisfied. It also requires progressive liberalization of safeguards over the course of their application.

The Agreement on Safeguards appears to have achieved its central objective. Preexisting gray area measures have been abolished, and WTO members have not sought to introduce new gray area measures. Likewise, formal reliance on GATT Article XIX is now far more common than in the waning years of GATT.

Safeguards have been a frequent target of litigation in the WTO system, however, and in every case to date that has reached the point of a formal ruling, the challenged measure has been ruled illegal. In some instances, national governments using safeguards have ignored obvious requirements under the law, but the litigation has also revealed some fundamental problems with the treaty text and with the ability of WTO members to implement it faithfully. The most difficult issues arise in connection with the requirement that import surges be attributed to "unforeseen developments" and with the requirement of a causal linkage between import surges and serious injury or threatened injury. We consider these issues further after reviewing the state of economic thinking about the proper role of safeguards in the international trading system.

The Economic Function of Safeguards The standard economic case for free trade counsels against the use of safeguards. They create the deadweight costs of protection while delaying the redeployment of resources in declining industries to more productive uses. Only if safeguard measures were routinely coupled with compensatory trade concessions would there be any basis for hope that their direct welfare effects are reasonably benign, but compensation has been the exception rather than the rule in practice.

Why, then, do nations employ safeguards, and why does the WTO/GATT system permit them? The answers to these questions are not entirely clear, although they have sparked a considerable amount of political and academic commentary.

Perhaps the most often invoked arguments for safeguard measures are those recited in U.S. national legislation. Section 201 of the U.S. Trade Act of 1974 indicates that safeguards may have one of two objectives   to restore the competitiveness of a declining industry or to facilitate its orderly contraction. Both of these possible justifications are problematic.

The notion that safeguards can be used to "restore competitiveness" presupposes that governments can accurately identify, and will choose to protect, only those industries that can become competitive again. But politically organized industries may seek and obtain trade protection irrespective of the impact that

it is likely to have on long-term competitiveness because of the short-term profits that it brings. Moreover even if governments were able to identify appropriate candidates for assistance and would properly exclude poor candidates, protection is not necessarily the best way to provide such assistance. Direct loans or subsidies to the troubled industry are in theory superior to protection. Finally, and most important, government intervention to restore competitiveness is simply unnecessary in countries with reasonable access to private capital markets. Private lenders will finance efforts to become "competitive" as long as the returns from such investments justify the apparent risk. Their unwillingness to do so is a strong signal that the investment is not justified, and there will often be little basis for governments to second guess the judgment of the capital markets.

The suggestion that safeguards can facilitate "orderly contraction" is also suspect. Beyond question, safeguards can slow the rate of industrial contraction in response to import competition, or end contraction altogether. But such delay is ordinarily an economic vice, not a virtue. Absent some market failure, resources will move to alternative uses at an appropriate pace without safeguards. To be sure, market failures are possible, and an important possibility in this regard is downward wage stickiness, perhaps due to unionization or to subsidies to the unemployed. But trade protection is likely to be an inferior policy instrument for addressing an unemployment problem that results from wage stickiness. Direct intervention in the labor market, through employment or retraining subsidies, targets the wage stickiness problem without introducing the costs of protection. Moreover safeguards are temporary and may thus do little more than postpone the unemployment problem to a future period.

For these reasons, modern economic commentary emphasizes alternative explanations for the existence of safeguards in the WTO/GATT system, generally grounded in political economy considerations. These accounts fall into roughly three categories. The first is the notion that safeguards afford a domestic political "safety valve" for protectionist pressures, deflecting them into an administrative process that ultimately affords less protection than would be put in place by the legislature absent the safety valve (see Lawrence and Litan 1986). Whatever the merits of the safety valve hypothesis as a justification for an administrative route to protection in a domestic political system such as that of the United States, however, it affords a questionable basis for safeguards in the WTO system. If the putative policy goal is to create obstacles to domestic interest groups seeking new or renewed protection, inescapable international commitments to eschew trade protection would seem helpful, which would themselves allow legislatures to "deflect" protectionist pressures by insisting that their hands were tied.

A second political economy account of safeguards is that of Bagwell and Staiger (1990, 2002), who argue that the function of a safeguard mechanism in the WTO/GATT system is to legalize behavior that might otherwise count as "cheating" and cause the system to unravel. A temptation to "cheat" arises in their framework because nations are "large" and can use tariffs to improve their terms of trade. The desire to cheat on tariff commitments increases in the face of import surges because cheating then yields more tariff revenue. If cheating occurs, Bagwell and Staiger assume that nations will revert to their preferred tariff policies in the absence of cooperation. To avert this prospect, it may be in the interest of the parties to a trade agreement to permit temporary deviation from tariff commitments in response to temporary import surges.

Although Bagwell and Staiger's account of safeguards is clever and rigorous, it has some difficulty explaining the structure of safeguard rules in practice. Among other things, nothing in their theory makes the condition of the import-competing industry relevant to the use of safeguard measures—any import surge suffices for the temptation to "cheat" to arise. Yet, the requirement of "serious injury" or threat thereof to an import-competing industry is central to the prerequisites for safeguards under WTO law. In addition, it is not clear that safeguards are needed to address the problem of "cheating" in a system with a well-functioning dispute resolution

system and calibrated retaliation such as that of the WTO.

A third political economy account of safeguards, first suggested by Dam (1970) and elaborated by Sykes (1991, 2006), posits that safeguards reduce the political risk of trade concessions under conditions of uncertainty. The analysis builds on Hillman (1982) and Baldwin (1982), which explain why declining industries will invest more resources than other industries (ceteris paribus) in lobbying for protection. Because it may be difficult for trade negotiators to predict the trade impact of concessions on their constituent industries, they may discover ex post that an import-competing industry is more harmed by concessions than expected and is clamoring loudly for relief. Temporary protection for that industry while its sunk capital depreciates may ameliorate heavy domestic pressures for protection. Because trade negotiators know that they can deviate from the bargain under these circumstances, they may be more willing to grant concessions ex ante. The welfare effects of safeguards then depend on whether safeguards do more to facilitate additional trade concessions ex ante or to allow renewed protection ex post.

Legal Issues If safeguards have any useful role to play in the trading system, their ability to play that role going forward is in some jeopardy. As noted, every safeguard measure that has been challenged to date since the inception of the WTO, and that has reached the stage of a formal ruling, has been found to be illegal. The problem is not simply a matter of WTO members failing to comply with clear obligations. Rather, the treaty text as interpreted by the WTO Appellate Body contains some confusing and illogical requirements that members have great difficulty interpreting (see Sykes 2006).

GATT Article XIX requires a linkage between the import surge and "unforeseen developments." This obligation raises two challenging issues what "counts" as an "unforeseen development," and how do members demonstrate a linkage between it and the import surge? The difficulties here have led to findings of illegality in several cases. It remains unclear how members are to determine what is "foreseen" and whether every "unforeseen" shock that leads to a rise in imports is a permissible predicate for safeguards (consider an unforeseen domestic supply shock, for example, that precipitates an import surge). Further, once an appropriate unforeseen development has been identified, the task remains to show how that development has affected import volumes. To do so convincingly may require elaborate econometric modeling, an exercise that is difficult to undertake given the time and data limitations in safeguards cases.

Article XIX also requires a causal linkage between increased imports and injury. The interpretation of this requirement has proven problematic, among other reasons, because import quantities are endogenous. It is thus difficult to treat import quantities as a causal variable at all. Moreover the Agreement on Safeguards requires national authorities to ensure that injury attributable to "other factors" is not erroneously "attributed" to imports. The task of defining the "other factors" has proceeded quite incoherently, even to the point of deeming an underlying cause of increased imports to be an "other factor." It then becomes impossible to distinguish the causal effect of "imports" from the causal effect of "other factors." It remains unclear how members can undertake the requisite nonattribution analysis in a fashion that will withstand legal scrutiny.

The future role of safeguards in the WTO system is thus in some doubt. An obvious concern is that if nations cannot employ safeguards legally, they will drift back toward the extralegal measures that motivated the Agreement on Safeguards during the Uruguay Round, and the key achievements of that agreement may be undermined. As of 2007, the larger trading entities such as the United States and the European Union whose safeguards are almost certain to be subject to legal challenge have simply stopped using them. The long-term systemic effects of this state of affairs remain to be seen.

See also General Agreement on Tariffs and Trade (GATT); nontariff measures; political economy of trade policy; Uruguay Round; World Trade Organization

FURTHER READING

Bagwell, Kyle, and Robert W. Staiger. 1990. "A Theory of Managed Trade." *American Economic Review* 80 (4): 779 95. Looks at safeguards as an aid to sustained co operation in trade agreements.

———. 2002. *The Economics of the World Trading System*. Cambridge, MA: MIT Press. Develops terms of trade model of trade agreements and applies them to various topics, including safeguards.

Baldwin, Robert E. 1982. "The Political Economy of Pro tectionism." In *Import Competition and Response*, edited by Jagdish N. Bhagwati. Chicago: University of Chicago Press, 263 94. Discusses the pressures for protection from troubled industries.

Dam, Kenneth. 1970. *The GATT: Law and International Economic Organization*. Chicago: University of Chicago Press. Classic reference on economic structure of GATT.

Hillman, Arye L. 1982. "Declining Industries and Political Support Protectionist Motives." *American Economic Review* 72 (5): 1180 87. Discusses the pressures for protection from troubled industries.

Lawrence, Robert Z., and Robert E. Litan. 1986. *Saving Free Trade*. Washington, DC: Brookings Institution. De velops safety valve theory of safeguards.

Sykes, Alan O. 1991. "Protectionism as a 'Safeguard': A Positive Analysis of the GATT 'Escape Clause' with Normative Speculations." *University of Chicago Law Review* 58 (1): 255 305. Looks at safeguards as a re sponse to political uncertainty and a device for facil itating trade concessions.

———. 2006. *The WTO Agreement on Safeguards: A Commentary*. Oxford: Oxford University Press. A comprehensive review of law and economics safeguards in the WTO system.

ALAN O. SYKES

■ sanctions

International economic sanctions are often favored by nation-states and by international organizations as a means of projecting power or influencing another government's behavior without resorting to military conflict. The utility of sanctions as an instrument of foreign policy is attested to by their growing popu-larity since the end of the Cold War. Economic sanctions include trade sanctions, which restrict imports from or exports to a target country; investment sanctions, which include restrictions on capital flows to the target and, in some cases, mandatory disin-vestment; and more narrowly targeted so-called smart sanctions, such as asset freezes and travel bans on individual members of the target nation's ruling elite.

A common argument against sanctions is that they are ineffective tools of foreign policy because of the relative ease with which target countries can find al-ternative markets and suppliers. In a study of 116 sanctions episodes imposed since 1914, for example, Hufbauer, Schott, and Elliott (1990) concluded that only 34 percent were successful in achieving their political objectives. Even this estimate may be opti-mistic, however, especially if success is interpreted to mean that sanctions were the *primary* factor con-tributing to the desired outcome (Pape 1997). An alternative view is that sanctions are necessarily suc-cessful as long as they impose costs on the target's behavior or enhance the sanctioner's international reputation. Galtung (1967) was one of the first to point out that sanctions are often imposed not pri-marily for instrumental purposes that is, to induce the target to comply with the sanctioner's demands but instead for expressive or demonstrative purposes. Thus one reason for sanctions' inefficacy is that they may not be designed to be effective in the first place: economic sanctions that are costly to a target are also costly to the sanctioner and are therefore likely to be avoided in favor of less costly, sometimes merely symbolic, measures.

Economic and Political Effects of Sanctions Both the sanctioning nation(s) and the target nation are, in general, made worse off by a trade embargo. The degree to which trade sanctions reduce welfare in these nations depends on the sanctions' terms-of-trade effects, which are larger in the case of multi-lateral than unilateral sanctions. Moreover any price distortions caused by sanctions inevitably create opportunities for nonsanctioning third parties to capture rents by continuing to trade with the target. Under multilateral sanctions, much sanctions-busting activity is likely to involve traders in the target

nation itself, thereby channeling some of the rents into the very country that is supposed to be punished (Kaempfer and Lowenberg 2007). Sanctions rents might even, perversely, enrich the target country's own rulers if they are able to participate in sanctions-busting trade. Empirical studies have confirmed that multilateral sanctions, despite, and indeed because of, their greater terms-of-trade effects, are typically less successful than unilateral sanctions.

Financial and capital sanctions, too, can have unintended consequences. In the case of disinvestment, the physical plant and capacity previously owned by foreigners is purchased by domestic capital owners at reduced prices, which increases their rate of return. The resulting windfall gain to domestic capital owners potentially increases the tax base available to the government to finance its objectionable policy (Kaempfer and Lowenberg 2007). Of course, in the long run, a decrease in the inflow of new capital and technology from abroad is likely to limit the target's growth potential.

Even if sanctions impose substantial economic damage on a target nation, there is no guarantee that the target will reform its behavior in conformity with the sanctioner's demands. In his classic study, Galtung (1967) noted that sanctions are often followed by increased levels of political integration in the target country, the so-called rally-around-the-flag effect, which has since captured the attention of many observers.

Sanctions have distributional effects. For example, an embargo on exports to a target country benefits domestic producers of importables in the target country at the expense of consumers. There are similar distributional effects within the sanctioning country. It follows that interest group politics in both countries are important in explaining the political origins and outcomes of sanctions. Thus the level of sanctions imposed by a sanctioning country depends on the relative political influences of prosanctions and antisanctions groups within that nation's domestic polity. In the target state, sanctions that are successful in bringing about a change in policy in the desired direction will generally do so by diminishing the political influence of the regime's supporters

relative to that of its opponents (Kaempfer and Lowenberg 2007). There is considerable consensus on the utility of smart sanctions that pinpoint those groups most responsible for the objectionable policy while minimizing "collateral damage" to innocent citizens (Cortright and Lopez 2002).

Game-theoretic work on sanctions demonstrates that, given perfect information, sanctions would never be implemented: if a threatened sanction were sufficiently effective, the target would comply immediately, obviating the need to impose the sanction, while if the sanction were ineffective, the sanctioner would not threaten it in the first place (Eaton and Engers 1992). Thus the sanctions that are most likely to succeed will often do so as threats, without having to be imposed at all. It has also been shown that the success of sanctions depends on conflict expectations: target states that anticipate future conflict with a sanctioner are relatively unlikely to comply with the sanctioner's demands, although these are precisely the countries that are most likely to be sanctioned (Drezner 1998).

Empirical Findings on Sanctions Many studies have attempted to identify the correlates of sanctions success. The most common findings are that the success of sanctions is positively correlated with political instability and economic weakness in the target country and with close, cordial ties between sanctioner and target prior to sanctions. Some scholars also find a significant positive relationship between the cost of the sanctions to the target, measured as a percentage of the target's gross national product, and the success of the sanctions.

Most of the empirical literature uses the data compiled by Hufbauer, Schott, and Elliott (1990), which include only those sanctions that were actually applied, despite the fact that successful sanctions often end at the threat stage. Consequently empirical studies of actually implemented sanctions are likely to be biased against sanctions success. Such selection bias is part of a broader problem, namely, that the factors determining whether sanctions are used are intrinsically linked to the factors determining their success. Recent empirical work therefore uses simultaneous equations techniques to estimate models

of jointly determined policy choice and sanctions outcome.

Another focus of empirical inquiry is the role of political regime type in determining sanctions usage and success. The interest of sanctions scholars in regime type stems from the literature on the democratic peace, which is the theory that democracies typically do not go to war against other democracies. Lektzian and Souva (2003) investigate whether there is an analogous "economic peace" among democracies. They find that, although democracies impose sanctions more often than other regime types, democracies are indeed less likely to sanction other democracies than they are to sanction nondemocracies. However, others have found evidence that sanctions are more successful when imposed against democracies than nondemocracies, suggesting that democracies make more attractive targets (Nooruddin 2002).

See also nontariff measures; terms of trade

FURTHER READING

Cortright, David, and George A. Lopez, eds. 2002. *Smart Sanctions: Targeting Economic Statecraft*. Lanham, MD: Rowman and Littlefield. Contains several illustrative case studies of smart sanctions.

Drezner, Daniel W. 1998. "Conflict Expectations and the Paradox of Economic Coercion." *International Studies Quarterly* 42 (4): 709 31. Emphasizes the importance of future conflict expectations in determining sanctions initiation and compliance.

Eaton, Jonathan, and Maxim Engers. 1992. "Sanctions." *Journal of Political Economy* 100 (5): 899 928. One of the first and most often cited game theoretic treatments of sanctions.

Galtung, Johan. 1967. "On the Effects of International Economic Sanctions: With Examples from the Case of Rhodesia." *World Politics* 19 (3): 378 416. The earliest classic in the international relations literature on sanctions, considered a seminal work and widely cited.

Hufbauer, Gary Clyde, Jeffrey J. Schott, and Kimberly Ann Elliott. 1990. *Economic Sanctions Reconsidered: History and Current Policy*. 2d ed. Washington, DC: Institute for International Economics. The main source of data for empirical work on sanctions.

Kaempfer, William H., and Anton D. Lowenberg. 2007. "The Political Economy of Economic Sanctions." In *Handbook of Defense Economics*, vol. 2, *Defense in a Globalized World*, edited by Todd Sandler and Keith Hartley. Amsterdam: Elsevier North Holland, 867 911. Provides a wide ranging survey of the sanctions literature and an exposition of the interest group model originally developed by Kaempfer and Lowenberg in a September 1988 *American Economic Review* article.

Lektzian, David, and Mark Souva. 2003. "The Economic Peace between Democracies: Economic Sanctions and Domestic Institutions." *Journal of Peace Research* 40 (6): 641 60. The first attempt to extend the democratic peace literature to sanctions.

Nooruddin, Irfan. 2002. "Modeling Selection Bias in Studies of Sanctions Efficacy." *International Interactions* 28 (1): 59 75. A good example of recent empirical work that addresses the problem of selection bias in sanctions data.

Pape, Robert A. 1997. "Why Economic Sanctions Do Not Work." *International Security* 22 (2): 90 136. An influential critique of sanctions as a policy instrument.

WILLIAM H. KAEMPFER
AND ANTON D. LOWENBERG

■ sanitary and phytosanitary measures

Since the 1990s, attention to agricultural trade policy has widened to include issues related to technical barriers to trade, particularly sanitary and phytosanitary (SPS) measures that regulate movement of products across international borders. SPS measures protect plant, animal, and human health from hazards such as pests, diseases, contaminants, and toxins that might be found in imported agricultural and forestry products. Unlike most nontariff measures, they are potentially welfare increasing because they may correct market failures stemming from externalities that arise from importing goods that may be accompanied by these hazards. But like other regulations, they arise from an equilibrium process of policy determination. The observed levels of intervention emerge from the interaction of the demands for measures by various domestic interest groups (including producers, processing industries, and

consumers) and the supply of these barriers by policymakers. That regulatory processes can be captured by interest groups with a vested interest in a particular regulatory outcome is well recognized in the economic theory of regulation, and cases involving the promulgation of SPS measures are no exception. Since the 1990s, objective economic analysis and international trade rules have provided the basis for reform of some SPS regulations that lowered net social welfare by restricting imports that posed negligible risks.

Measurement of Economic Effects Although traditional trade barriers have been extensively studied, informing policy trade-offs and negotiating priorities in the Uruguay and Doha trade rounds, far less is known about the impacts of SPS measures on trade. The analytical complexity and scant data available for quantifying the economic effects of SPS measures pose significant challenges for analysts. Conceptual difficulties stem from the heterogeneous nature of SPS measures, the fact that multiple measures are aimed at a particular product, and that measures are difficult to aggregate.

Data scarcity is perhaps an even more significant challenge. No registry for SPS measures comparable to tariff schedules exists at either national or international levels. The UN Conference on Trade and Development is leading an effort to update its Trade Analysis and Information Systems database, an international inventory of nontariff barriers, to fill this data gap. At present, however, direct measures of the economic effects of SPS and other technical regulations are limited to the evidence from business surveys. Surveys by the Organisation for Economic Cooperation and Development and the World Bank indicate that these regulations can significantly increase costs or limit market access for exporters of selected products in developing and developed countries alike.

Indirect methods of measuring the effects of SPS measures on trade are important because of the paucity of direct measures. These methods rely on comparison of prices or are inferred from trade quantities in the context of a well-specified model of trade flows. Many studies in the literature have used partial equilibrium simulation models or gravity models to estimate the costs of specific measures, including Australia's ban on banana imports, the European Union's (EU's) ban on growth hormones in beef production, Japan's restrictive fire blight measures for apple imports, and the EU's aflatoxin standards for dried nuts and fruits. A smaller number of case studies have integrated risk and economic assessment in the analysis of SPS measures under alternative risk scenarios to inform specific policy changes for example, the gradual lifting of the United States' long-standing ban on imports of Mexican avocados. These studies have made important contributions to the development of methodology and policy formulation, but their results are not easily generalizable to other products or sectors in other countries.

More recently, a few econometric studies have examined the effects of SPS and other nontariff measures on food trade at a more aggregate level, finding that their effects are larger than tariff effects for product categories, including fruits and vegetables as well as meats. This emerging body of literature, which is still in its early stages and fraught with measurement, endogeneity, and other econometric difficulties, reinforces the perception that SPS barriers have a substantial influence on agricultural trade.

International Trade Rules Although it has been difficult to assess the aggregate effects of SPS regulations on trade in agricultural goods, or to evaluate their relative importance in the world trading system, there has long been broad recognition that these measures can significantly impede trade. Disciplines on SPS measures and other technical regulations were included in the original General Agreement on Tariffs and Trade (GATT) adopted by the Contracting Parties in 1947. The GATT treaty recognized the need to subject domestic regulations to international scrutiny so that the strategic application of these measures would not subvert the commercial opportunities created by other trade policy reforms.

The original GATT rules for SPS and other technical measures were expanded in the Tokyo

Round (1973 79) of multilateral trade negotiations and again during the Uruguay Round (1986 93) when a separate agreement for SPS measures was established. The challenge before the negotiators of the World Trade Organization (WTO) Agreement on the Application of Sanitary and Phytosanitary Measures (the SPS Agreement) was to create a set of rules that would strike the proper balance between allowing health and environmental protection while disallowing mercantilist regulatory protectionism. In broad terms, the SPS Agreement recognizes the right of each WTO member to adopt measures that achieve its "appropriate level of protection" but requires such measures to be based on a scientific assessment of the risks and to be applied only to the extent necessary to achieve the country's public health or environmental goals. As a benchmark, the SPS Agreement recognizes the international standards promulgated by the *Codex Alimentarius*, the International Organization of Epizootics, or the International Plant Protection Convention to be " safe harbor" standards. Members adopting these standards are " rebuttably presumed" to be in compliance with the agreement.

The compliance of countries with the SPS Agreement is reinforced by the WTO's formal dispute settlement procedures. Only a few conflicts over SPS measures have led to the establishment of dispute panels, but these few cases have played a critical role in establishing how trade rules place bounds on national sovereignty in this policy area. Of 30 requests for formal consultations on SPS measures from 1995 to 2006, 11 have advanced to dispute settlement proceedings. Landmark SPS disputes included the EU ban on hormone-treated beef and the EU's regulations for approval and marketing of biotech products.

The EU's defense of its hormone ban rested on its claims that existing international standards for hormone use in beef cattle did not meet its public health goals, and that the ban represented a precautionary approach to managing uncertain risks. In their rulings, the WTO affirmed the right of WTO members to establish a level of consumer protection that was higher than international health standards. The ban

was nonetheless judged to be in violation of the SPS Agreement because it was not backed by a scientific risk assessment. The WTO likewise ruled against the EU and some individual member states in the more recent biotech case because risk assessments did not support the disputed measures. These results remove a degree of national political sovereignty for regulations in cases in which evidence has not been marshaled to demonstrate any risk from trade. But the limits to what the WTO can achieve in this policy area is seen in the fact that the formal rulings have triggered additional scientific assessments rather than meaningful policy changes leading to increased trade in these two long-running disputes.

Prospects for Regulatory Reform of SPS Measures Although current international trade rules curb the protectionist abuse of SPS measures, they still allow countries to adopt measures for which global or even national costs outweigh their national benefits. Economic analysis that examines the benefits (reduced risks) and costs (reduced trade) of SPS regulations is important for regulatory reform efforts that seek to identify the most efficient policy instruments to manage SPS risks. Failure to work through the challenges of data and methodology will create the potential for anecdote and allegation to play a larger role in SPS policy formulation, which would likely result in increased costs to consumers and increased tensions between trading partners.

See also gravity models; nontariff measures; partial equilibrium models; technical barriers to trade; World Trade Organization dispute settlement

FURTHER READING

Beghin, J. C., and J. C. Bureau. 2002. "Quantitative Policy Analysis of Sanitary, Phytosanitary, and Technical Barriers to Trade." *Economie Internationale* 87: 107 30. A concise description and evaluation of the various methods available for measuring the effects of SPS and other technical measures on trade and welfare.

Josling, T., D. Roberts, and D. Orden. 2004. *Food Regulation and Trade: Toward a Safe and Open Global System*. Washington, DC: Institute for International Economics. A roadmap of SPS and other technical measures

in international food markets, including definitions and classification, explanation of economic rationales and effects, and authoritative description and analysis of the institutions that govern the intersection of regulation and trade.

Maskus, Keith E., and John S. Wilson. 2001. *Quantifying the Impact of Technical Barriers to Trade: Can It Be Done?* Ann Arbor: University of Michigan Press. An often cited early survey of methods for empirical analysis of SPS and other regulations on trade.

DONNA ROBERTS

■ savings glut

See global imbalances; twin deficits

■ seigniorage

Seigniorage is profit from money creation, a way for governments to generate revenue without levying conventional taxes. In the days of commodity money, seigniorage revenue was the difference between the face value of the minted coins and the actual market value of the precious metal they contained. When this markup was insufficient for the government's revenue needs, the authorities might substitute less valuable base metal for some of the precious metal that was supposed to be in the coins. Such a practice has a long history, dating back at least to Roman times. Although this allowed the government to issue more coins without acquiring more precious metal, the coins quickly depreciated as agents became aware of their less valuable content. The continued partial precious metal backing still placed some limit on the possible issuance.

With the demise of the gold standard in the early 1930s, almost all nations abandoned commodity-backing for their currencies and adopted a fiat money standard under which the paper issued is backed by nothing more than faith and confidence in the issuer. Under this system, the cost of issuance declines close to zero and there is no longer any limit on the quantity of money that can be issued. Under a fiat

money system, seigniorage revenue is given by the product of the inflation rate and the inflation tax base. This inflation tax base reflects the purchasing power of the public's money holdings and is the level of *real* money balances (nominal money holdings divided by the price level). Undertaking more rapid monetary expansion causes the inflation rate to rise, but the revenue effects are partially offset as individuals attempt to quickly spend the extra money before it depreciates further. If people spend money faster than it is being printed, the rate of price increase comes to exceed the rate of money issuance.

Hyperinflation A government that is unable to fund its expenditures through conventional taxes or bond sales may become dependent on seigniorage revenues to maintain its existence. Attempts to raise seigniorage revenues are, however, not only inflationary but eventually self-defeating. Under circumstances where the decline in real money balances becomes proportionately larger than the rise in the inflation rate, the inflationary policy actually backfires and lowers seigniorage revenue. The economist Phillip Cagan's classic analysis of hyperinflations, with inflation rates exceeding 50 percent per month, suggested instances where the process had, in fact, been pushed beyond the revenue-maximizing point. It is possible that lags in the adjustment of inflation expectations may actually have allowed continued seigniorage gains, however, and subsequent analysis of the 1921 23 German hyperinflation points to seigniorage levels rising year by year (see Cukierman 1988).

Whatever success the German government may have attained in raising seigniorage revenues, such surging, and highly volatile, inflation rates interfere with the price mechanism and cloud production and employment decisions. Meanwhile attempts to economize on money balances require devoting more and more time to simply turning over the currency. By the end of the German hyperinflation, workers were being paid more than once a day because otherwise the value of the money would fall too much between the beginning and end of the shift! Other innovations, such as the indexation of bank deposits to inflation or establishing deposits de-

nominated in a stable-value currency, require even faster rates of note issue to maintain seigniorage revenues, given that part of the money supply is now insulated from the inflation tax. This factor was particularly evident in the record-breaking post World War II Hungarian hyperinflation. Negative effects on economic growth emerge well before such extreme circumstances, however. Recent work suggests that significant negative effects of inflation may emerge below the 10 percent level in both industrial and developing economies (Burdekin et al. 2004).

Certainly, even though conventional taxes impose their own distortions on the economy, few would argue that this justifies reliance on seigniorage as a deliberate policy choice. Dependence on seigniorage revenue seems, in practice, to be highly correlated with the degree of political instability. Vulnerability to this effect tends to be greatest in developing economies that are less democratic and/or more socially polarized (Aisen and Veiga 2005). High indebtedness is another common factor. In this respect, even a government that is initially able to fund its deficit through selling bonds to the public may eventually resort to inflation finance as the amount of debt approaches the saturation point. Thomas Sargent and Neil Wallace argue that such "unpleasant monetarist arithmetic" implies that, in the face of continued budget deficits, tighter monetary policy today simply implies more inflationary policy tomorrow once the limit on bond issuance is reached (see Sargent 1993). Furthermore the greater the outstanding stock of bonds, the greater the potential governmental gains from inflating away the real value of these obligations through inflation which, in turn, likely limits the demand for such bonds unless inflation protection is built in.

The typical mechanism of inflation finance is for the government to sell bonds to the central bank, which then immediately "monetizes" the debt with new money emissions. This is therefore properly characterized as money finance rather than true bond finance and provides the government with new funds to spend only via excess money creation. Conversely selling bonds to the public, rather than the central bank, is likely to be far less inflationary and has no direct effect on the money supply.

Making the central bank independent of the fiscal authority would seem to be a way of ending automatic government access to the printing press, and nations with independent central banks have historically tended to have lower inflation rates. But even in the United States, large deficits such as those incurred under the Reagan administration put considerable pressure on the central bank to at least partially monetize them insofar as very large bond issues push down bond prices and put upward pressure on interest rates. Moreover in a developing economy where the market for government bonds is much thinner, it is unrealistic to expect the central bank, nominally independent or not, to resist monetization pressures when no other financing options exist. Instituting central bank independence is probably better thought of as a way of discouraging future fiscal profligacy rather than something that alone can end an ongoing reliance on deficit monetization and seigniorage.

Reducing Reliance on Seigniorage Revenue A more drastic way of weaning a government away from reliance on seigniorage revenue is to dollarize the economy, following the example of countries like Panama (since 1904) and Ecuador (since 2000), which abandoned domestic currency issuance and adopted the U.S. dollar as their monetary standard. Another option is to maintain the domestic currency but link it to the U.S. dollar or another currency via a currency board arrangement, whereby the monetary authority commits to exchange local currency for the foreign currency on demand at a predetermined, fixed rate of exchange. Such a strategy still allows for some seigniorage revenue on interest-bearing dollar-denominated assets held to back the local currency. Although Hong Kong's currency board arrangement with the U.S. dollar has been maintained since 1983, Argentina's link with the U.S. dollar collapsed in 2001 in the midst of soaring unemployment and political unrest. The Argentinean case not only illustrates the potential dangers of tying the domestic currency to the dollar but also serves as a reminder that no such announced commitment is truly

irrevocable. Another alternative is to enter into a currency union, such as the euro zone, that eliminates the scope for domestic inflation finance by dispensing with the national currency entirely yet allows participating nations to share in seigniorage revenue earned by the group as a whole.

See also commodity-price pegging; common currency; currency board arrangement (CBA); currency crisis; currency substitution and dollarization; dollar standard; euro; European Central Bank; Federal Reserve Board; gold standard, international; money supply; multiple currencies; quantity theory of money

FURTHER READING

Aisen, Ari, and Francisco José Veiga. 2005. "The Political Economy of Seigniorage." IMF Working Paper 05/175. Washington, DC: International Monetary Fund. Available at http://www.imf.org/external/pubs/ft/wp/2005/wp05175.pdf. Reviews factors influencing the extent to which governments rely on seigniorage revenue and presents evidence on the importance of political instability.

Banaian, King, J. Harold McClure, and Thomas D. Willett. 1994. "The Inflation Tax Is Likely to Be Inefficient at Any Level." *Kredit und Kapital* 27 (1): 30 42. Emphasizes that the costs of resorting to the inflation tax are significantly augmented by the increased uncertainty that typically accompanies higher levels of inflation.

Burdekin, Richard C. K., Arthur T. Denzau, Manfred W. Keil, Thitithep Sitthiyot, and Thomas D. Willett. 2004. "When Does Inflation Hurt Economic Growth? Different Nonlinearities for Different Economies." *Journal of Macroeconomics* 26 (3): 519 32. Examines the thresholds at which inflation may start to exert significant negative effects on growth in both industrialized and developing economies.

Cagan, Phillip. 1956. "The Monetary Dynamics of Hyperinflation." In *Studies in the Quantity Theory of Money*, edited by Milton Friedman. Chicago: University of Chicago Press, 25 117. Classic analysis of money demand and seigniorage using data from a number of post World War I and post World War II experiences. Cagan's work essentially gave birth to the study of hyperinflation as a burgeoning subfield in macroeconomics.

Cukierman, Alex. 1988. "Rapid Inflation Deliberate Policy of Miscalculation?" *Carnegie Rochester Conference Series on Public Policy* 29, edited by K. Brunner and B. McCallum (autumn 1988), 11 84. Reprinted in P. Siklos, ed. 1995. *Great Inflations of the 20th Century*. Aldershot, UK: Edward Elgar. Reexamines the seigniorage revenue raised during the post World War I German hyperinflation and suggests that the inflation may have been more profitable to the government than the original Cagan analysis implied.

Gros, Daniel. 2004. "Profiting from the Euro? Seigniorage Gains from Euro Area Accession." *Journal of Common Market Studies* 42 (4): 795 813. Examines the distribution of the European Central Bank's seigniorage revenue and the prospective gains accruing to the new member states from Central and Eastern Europe.

Sargent, Thomas J. 1993. *Rational Expectations and Inflation*. 2d ed. New York: HarperCollins. Chapter 5 covers the "unpleasant monetarist arithmetic" when tight monetary policy is combined with continuing budget deficits, and chapters 2 and 6 assess the implications of the Reagan deficits and their possible rationale.

RICHARD C. K. BURDEKIN

■ sequencing of financial sector reform

Sequencing of financial sector reform is concerned with the ordering of reforms in the financial system and with designing a path for these reforms that will improve economic welfare. Economists generally acknowledge that liberalized or market-based financial systems enhance the mobilization and allocation of financial resources and, therefore, economic growth and welfare more than repressed financial systems in which allocation of financial resources is subject to administration. History has shown, however, that when a country moves from a repressed to a market-based system, it risks its financial and economic stability and can face an economic and financial crisis. Liberalization of certain financial transactions before the necessary financial infrastructure is in place can contribute to the incidence of these financial crises. The sequencing of the financial sector reforms is intended to ensure an orderly

transition from repressed to market-based financial systems that avoids serious economic and financial crises.

Policy Issues Surrounding Sequencing The earliest literature on sequencing financial sector reform addressed the reform of financial systems as part of broader programs of economic reform that included, for example, trade liberalization and labor market reforms. The prevailing belief was that financial liberalization should occur late in the reform process since financial sector variables were seen as responding to relative prices and savings and investment decisions. Liberalizing financial systems before successful economic reforms that set the "correct" relative prices would lead to a less-efficient allocation of financial resources and could be destabilizing, as observed among some countries in Latin America during the 1970s.

A second strand of research addressed the speed of reform: whether it would be better to follow a gradualist approach or to have a rapid "big bang" program that reformed several sectors simultaneously. The arguments for "big bang" reform programs were motivated largely by political economy considerations, especially the need for a critical mass of reforms to break entrenched inertia and vested interests and to provide domestic and international credibility to the reform program. Financial sector reforms were seen as a key component of the "big bang" reform programs since they were often easier to implement than other components, such as labor market reform.

A third strand of the work on sequencing has focused on the design of the programs of financial sector reform. This strand identifies the major risks as well as benefits of the financial sector reform and seeks to mitigate these risks through the design of the financial liberalization. The emphasis is not on the speed of the reform but instead on putting in place a consistent set of polices that takes into account each country's unique circumstances.

A fourth related strand has focused on the liberalization of the capital account and, specifically, on the question of when countries should liberalize capital flows. This issue received additional attention following the Asian financial crisis of the mid-1990s, which was viewed as the result partly of a premature or an inappropriately sequenced liberalization of short-term capital flows.

Key Policy Insights As countries reform their financial systems, it is essential that they have instruments that allow them to exercise effective monetary control in a liberalized environment. The liberalization of financial systems is typically associated with significant initial expansions in monetary and credit aggregates and capital inflows. The expansions are largely the result of portfolio adjustments to the liberalized financial systems, but they can also have potentially destabilizing macroeconomic effects through increasing inflation or weakening of the balance of payments. Thus the authorities need to have flexible monetary instruments to manage these effects and avoid the buildup of potential fragility in the financial system that can result in financial crises. Typically, this has involved adopting techniques of indirect monetary control and following a consistent mix of monetary and exchange rate policies. For example, if countries seek to follow a restrictive interest rate policy to limit the credit expansions, they may need to adopt a flexible exchange rate regime.

Financial institutions typically need to be restructured early in the process of financial sector reform. Many of the financial crises that followed financial liberalizations can be traced to imprudent lending by weak banks. The bank restructuring programs include elements such as writing down the value of bad and doubtful debts and recapitalizing the banks to achieve a minimum level of capital at the beginning of the reform programs; reorganizing bank management and developing skills in risk assessment and management; introducing regulations of banks to limit large credit and currency risk exposures; and developing an effective system of prudential oversight. Experience has shown that banking failures can be costly to economies: they result in a loss of confidence in the currency and require significant government interventions to clean up bank balance sheets.

Implementation of the monetary and prudential elements of the reforms normally requires giving

some urgency to strengthening official institutions specifically, the central bank and the supervisory authorities   in the early stages of the financial sector reforms.

Reforms to develop longer-term capital markets and nonbank financial intermediaries have tended to occur at a later stage of financial sector reform. It usually is easier to develop the markets in longer-term instruments, the bond and securities markets, after those in short-term instruments are in place. Reforms of auditing and accounting, legal frameworks, bankruptcy, and payments and settlement systems generally should accompany these reforms. Derivative financial markets are normally the last to emerge since they rely on well-functioning markets in the underlying instruments on which the derivatives are based.

Payments and transfers on current international transactions have conventionally been liberalized ahead of capital movements. The former are important elements in the growth of trade in goods and services. The elimination of restrictions on current international transactions has often depended on foreign exchange market reforms.

A more controversial issue has been how to sequence the liberalization of capital movements with the domestic financial-sector reforms. One of the lessons of the Asian financial crises in the mid-1990s is that capital account liberalization should be approached cautiously with special attention to avoiding reliance on short-term capital flows. These crises were attributed in part to a premature opening of the capital account to short-term capital inflows. These flows were potentially destabilizing since, in the event of a loss of confidence, they could be withdrawn quickly, resulting in a balance of payments crisis. At the same time, capital movements can be beneficial as part of the financial-sector reform program. For example, allowing foreign direct investment (FDI) in financial institutions can help in their restructuring and in strengthening local technical and management skills. Notwithstanding the well-founded concerns about an overreliance on short-term capital movements, liberalization of certain short-term capital transactions is a key element in developing

efficient foreign exchange and domestic money markets. Hence capital account liberalization involves various complex elements, making it difficult to generalize on the optimal approaches.

Although it is neither desirable nor possible to have a one-size-fits-all approach to sequencing, some important technical and policy lessons have emerged on how to sequence reform measures to avoid major pitfalls. A key lesson is that at the beginning of the reform program, the authorities need to develop monetary instruments to maintain macroeconomic control. They will also probably need to restructure or recapitalize their commercial banks to avoid a major failure that could disrupt the domestic financial system and lead to a loss of international confidence. Finally they should approach the liberalization of short-term capital movements cautiously to avoid an overreliance on this source of capital.

See also banking crisis; capital controls; capital mobility; currency crisis; exchange rate regimes; financial crisis; financial liberalization; financial repression; hot money and sudden stops; impossible trinity

FURTHER READING

Harwood, Alison, and Bruce L. R. Smith, eds. 1997. *Sequencing? Financial Strategies for Developing Countries.* Washington, DC: Brookings Institution Press.

Johnston, R. Barry, and V. Sundararajan, eds. 1999. *Sequencing Financial Sector Reforms: Country Experience and Issues.* Washington, DC: International Monetary Fund.

R. BARRY JOHNSTON

■ services

See trade in services

■ shipping

Shipping, the movement of people and commodities by ship, has long been the dominant form of transportation for the movement of bulk commodities, especially over long distances. Many of the major

cities in the world have grown because of their sea-ports or access to good inland navigation. Physically, some 90 percent of world trade (nearly 7 billion tons) is carried by the international shipping industry, although because of the nature of the goods moved, the proportion is considerably less in terms of value. Shipping also is a growth industry, with the amount of seaborne trade in terms of tonne-miles (the amount carried, measured in metric tonnes, multiplied by the distance moved) rising from less than 6,000 in 1965 to nearly 30,000 in 2005.

As of 2006 there were about 50,000 mercantile vessels in the world, totaling some 650 million gross tons, engaged in international trade, and many more providing domestic coastal and ferry services. The merchant fleet consisted mainly of general cargo vessels (38 percent of the ships), oil tankers (25 percent), and dry bulk carriers (14 percent). It is, however, the container fleet that has expanded most rapidly, not only in relative numbers but also in size the average capacity rising from 1,944 twenty-foot equivalent units (TEUs) in 2002 to 2,108 in 2004. It is also a newer fleet: the average age of the overall merchant fleet is 12.5 years, but only 9.2 years for container ships. Passenger ships accounted for about 7 percent of the fleet. The majority of vessels register in developed countries (27 percent) or in the major open-registry ("flag-of-convenience") countries that allow foreign ships to register under their flag and offer attractive financial and other incentives to shipping companies to do so (Panama, Liberia, and the Bahamas, which register a total of 250 million gross tons). More than 97 percent of the world's merchant fleet was registered in just 35 countries.

The world's shipbuilding industry delivered 2,129 new ships, amounting to a record 70.5 million tons of new shipping capacity, in 2005, up from 49.4 million tons in 2004. This increased global net capacity by 7.2 percent. The major shipbuilding nations include South Korea, Japan, China, Germany, and Russia, with China taking an increasingly important role. Structurally, the shipbuilding industry tends to be more fragmented in Europe than in Asia where it is concentrated in a limited number of countries.

Although, because of technological advances, shipping is becoming more capital intensive and crews levels are getting smaller for any given size vessel, its overall size and growth rate means that it remains a major employer of labor. There are 466,000 officers and 721,000 sailors serving merchant ships that trade internationally with Organisation of Economic Co-operation and Development (OECD) countries. East Asia provides the majority of merchant ships' officers and developing countries the majority of sailors. Indirectly, the shipping industry creates much greater employment than these numbers indicate because ships need labor for their building, maintenance, chandelling (fitting out vessels), and vittling (supplying food for vessels).

Although there has been volatility over time, the costs of shipping have never been cheap and have varied considerably across markets. The cost of international freight transportation, for example, was 5.8 percent of the import value for developed countries in 2002 but ranged from 8.8 percent for developing nations overall to 12.4 percent for African countries. Although estimates of the income earned from international shipping are imprecise, the approximate figure is about $380 billion annually.

Structure of the Shipping Industry The shipping industry is segmented in a number of ways. The largest part of the industry involves the movement of cargo, with more limited transportation of people by ferries and cruise ships. A small number of passenger liners still provide scheduled services on oceanic routes, but this type of traffic has largely transferred to air transportation. The cargo part of the industry is divided into coastal activities and deep-sea shipping. The former serves point-to-point short-haul routes and acts as a distributor for large consolidating ports that collect and dispatch larger ships on deep-sea routes. Coastal shipping serves a key function in many countries where, for economic or geographical reasons, land transportation is of poor quality. For example, more than 70 percent of maritime activity in China is cabotage (coastal water transportation) traffic and accounts for nearly 55 percent of the total domestic ton-miles transported in the country.

Sliced another way, shipping offers regular, scheduled services ("liner services") and irregular sailings ("tramp shipping"). Liner services cater to customers who have partial shiploads often involving container shipments and who need to incorporate their cargo into a predetermined supply chain. Modern, just-in-time production demands reliability as well as speed in maritime logistics. To improve service, liner operations generally involve some form of cooperation between shipping companies to ensure regular sailings offering common-carrier services at frequent intervals. Traditionally, since the 1870s, when steam power made for reliable schedules with published rates, "conferences" have offered this service, although these have become refined into more integrated "consortia" and "alliances" that, for example, involve operating agreements allowing liners to share space on one another's ships. Containerization has provided scope for this interoperability by which ships from any line can carry standard containers. Although these agreements among shipping companies offer benefits in terms of coordinated, more stable rates and more frequent service, they also raise concerns about the concentration of market power that they may foster. Since the early 1990s, partly because of the wider use of long-term leases, greater competition for high-value cargo from air transportation, and tighter regulatory controls, but also because consolidation of liner operations has taken place, the number of conference type arrangements has declined.

Tramp shipping companies are often as large as their liner counterparts but are usually hired to carry entire shiploads of cargo, especially dry bulk or liquids. They often operate specialized vessels. In many cases these markets may involve the long-term leasing of vessels to a single customer, especially if they are for specific cargoes over a single route. The market for tramp shipping is highly competitive both for spot hires and for long-term leases. Rates tend to be highly volatile, reflecting the prevailing demand in the markets being served combined with the high degree of rigidity in short-term capacity. The rates are, at a macro level, highly correlated with international business cycles.

Economics of the Industry Shipping forms part of a more extensive supply chain. The industry directly interacts with ports, canals, and navigation systems and less directly with the operational surface modes (road, rail, and inland waterways) that service ports. This requires coordination of technology and of institutional structures if seamless service is to be provided to users, and so the industry has always been at the forefront of adopting up-to-date communications technology such as global positioning systems.

On the supply side, the technical peculiarities of the industry mean there are considerable economies of scale involved in the physical provision of individual ships. Simply put, a vessel 150 feet long, 20 feet high, and 20 feet wide has a capacity of 60,000 cubic feet and, assuming it is fully decked, has 12,800 square feet of steel plating. In contrast, a vessel 300 feet long, 40 feet high, and 40 feet wide has a carrying capacity of 480,000 cubic feet but requires only 51,200 square feet of plating. A doubling of dimensions increases capacity eightfold but the material needed to build the vessel rises only fourfold.

As engine technology and maritime engineering have allowed, the trend has been for larger cargo and passenger vessels for example, super-post Panamax (50,000 to 79,999 deadweight tons [dwt]) and Cape-size (more than 80,000 dwt) vessels used in the international movement of dry-bulk, ULCC (more than 300,000 dwt) crude oil tankers, and mega cruise liners. This challenges ports that do not have the depth or docking capabilities to handle such ships and water transit facilities, such as the Panama Canal, that do not have the width. Container traffic has grown considerably, especially in developing countries, and for this to be fully efficient ports also require specialized handling equipment that has to be upgraded as vessels get larger.

Institutionally shipping has always been seen as a risky activity, involving relatively large sums of capital and the commitment to tie up inventories for extended periods. Historically the sector expanded during the Classical Greek period only with the emergence of primitive insurance and later, in the Italian Renaissance, as more sophisticated

banking developed. Today the insurance market at Lloyds of London and financial markets around the world provide the lubricant for a multibillion-dollar industry.

The aggregate demand for shipping services is strongly tied to the state of the international economy. The worldwide recession of the early 1980s led to marked downturns in demand, and growth slowed during the Asian financial crisis of the late 1990s. But there can be wide variations in specific geographical and commodity markets. The demand for oil tankers, for example, can vary considerably according to the price of crude. Economic expansion in India and China has spurred rapid growth in those particular markets, with accompanying high freight rates. There are also spatial imbalances in demand, with significant difference in the demand for outbound and inbound services. Many developing countries, for example, are major exporters of bulk ores but are recipients of high-value, less bulky manufacturers. Tankers usually return from refineries to oil-producing countries in ballast (that is, empty of cargo, carrying only ballast to provide stability and maneuverability). This situation often leads to ships serving an "interconnected" network rather than a simple back-and-forth, "connected" routing. The American slave trade of the 18th century, for example involved a triangular pattern of service, with vessels taking European manufactures to Africa, collecting slaves in Africa for transportation to America, and bringing raw materials from America such as cotton and sugar back to Europe. This type of complex network is more difficult for specialized fleets such as tankers.

International trade cycles, combined with the time it takes to build new capacity and the competitive nature of much of the industry, often mean that there is a mismatch between supply and demand. In particular, given the costs of mothballing vessels during downturns in the business cycle, there is a tendency for excess capacity to exist over the full cycle. Estimates by the United Nations suggest, for example, that globally there was a residual (excess) fleet capacity amounting to some 21.7 million dwt (2.6 percent of the world fleet) in 2002, falling to 10.3 million dwt (1.2 percent) in 2003.

Regulating Shipping Markets Institutionally, long-standing debates have focused on the appropriate economic regulation of shipping to ensure that overall logistics chains are not distorted and on the social regulation of shipping to ensure safety and security. The shipping industry is thus subject to a wide range of laws and agreements, but the size and diversity of the sector make analysis and subsequent policy formulation difficult. The International Maritime Organization, an arm of the United Nations, provides broad global oversight, although other agencies, such as the United Nations Conference on Trade and Development (UNCTAD), become involved in maritime affairs when shipping is seen as a tradable service in its own right. At the macroregional level, national groupings such as the European Union play a role in determining the structure of the industry within more limited spatial markets.

Many developing countries have felt disadvantaged because their trade has been carried on ships from wealthier nations. Seeking to safeguard the position of developing countries, UNCTAD initiated in 1974 a "40/40/20 rule" whereby the freight being moved is shared between ships from the developed countries (40 percent), the developing country (40 percent), and cross-traders from third states (up to 20 percent). Developing countries often do not have the maritime capacity to handle their share of traffic and, because far more tonnage leaves most primary goods producing developing countries than is moved to them, efficiency drops.

Most nations allow free movement of shipping through their territorial waters and access to their ports, although there are limitations on the types of activity that may be pursued. Fishing and the extraction of minerals within territorial waters tends to be highly restricted. Many counties also limit maritime trade between their own ports (cabotage), even if international trade is allowed. The 1920 Merchant Marine (Jones) Act, for example, permits only U.S.-owned, -registered, and -built vessels, manned by citizen crews, to conduct cabotage between U.S. seaports.

Much of the economic regulation has focused on defining ownership and responsibilities for shipping and on ensuring that potential monopoly forces do not distort the market to the detriment of consignors or the public interest more broadly defined.

The growth of merchant shipping has always been vulnerable to attacks, and even in the early part of the 21st century there were more than 400 acts of piracy a year. Naval protection has thus often been seen as an important requisite of civilian shipping. Conversely merchant shipping has also been closely entwined with military operations, being a major part of many army and naval logistics operations. As a result, there are strong links between the merchant and military sides of shipping.

Despite the efforts of international bodies such as the World Trade Organization to remove tariffs and other institutional barriers to trade, the major factor that has facilitated the vast increase in the physical volume of goods trade has been technical improvements in the shipping sector. Without the advent of modern oil tankers, the energy required for production would not be available; without massive bulk carriers, food and raw materials could not be moved to places of production; and without the capabilities of container vessels, final products could not be moved to market. Shipping is the facilitator that allows the world economy to realize much of the division of labor so important for global economic growth.

See also air transportation; commodity chains; trade in services

FURTHER READING

International Ocean Institute and Dalhousie University Law School. Annual. *Ocean Year Book*. Chicago: University of Chicago Press. An annual publication initiated in 1978 that contains articles of current relevance to international shipping and related activities.

Maritime Policy and Management. Quarterly. London: Routledge. An international academic journal that contains peer reviewed articles on the shipping and port industries.

United Nations. Annual. *Review of Maritime Transport, Report by the UNCTAD Secretariat.* New York: United Nations. Contains information and data on trends in international shipping together with a number of in depth assessments of particular markets or trades.

KENNETH BUTTON

■ **Singapore issues**
See competition policy; government procurement; trade facilitation

■ **Smithsonian Agreement**

The Smithsonian Agreement was the understanding by which the Organisation for Economic Co-operation and Development (OECD) economies agreed to suspend the Bretton Woods system of fixed exchange rates in December 1971. The Bretton Woods conference had established a system of fixed exchange rates in 1944, encompassing nearly all the currencies of the free world, in which each currency was pegged to within a band of ±1 percent around a fixed parity against the U.S. dollar, and the dollar itself to a fixed price against gold. These systems had worked well and, with the exception of a few parity realignments in Europe in the 1960s, most countries had successfully kept to their parity values.

In the early 1960s, the United States still had a strong reserve position in which its gold holdings and foreign assets were larger than its liabilities to foreign institutions. But increasing balance of payment deficits later in the decade reduced those gold holdings, raised the United States' foreign liabilities, and weakened its reserve position. By the beginning of 1970, it was clear that the exchange rate regime was under threat when the U.S. balance of payments deficit sharply widened again and the (net) reserve position turned negative. In the first instance, these changes were the result of expansionary fiscal policies in Washington, driven mainly by President Lyndon Johnson's Great Society program and the expenditures of the Vietnam War. The result was a persistent tendency to large balance of payments deficits and a shortage of international liquidity, as the United States showed itself unwilling to restrain growth,

employment, and the expansionary pressures that were starting to increase inflation as well as aggregate demand. The initial response by foreign governments was to buy dollars and sell their own currencies in an effort to support the Bretton Woods system although some bought gold instead, which added to the American difficulties.

The United States meanwhile wanted to avoid devaluing the dollar (raising the dollar price of gold), not only to retain the support of the governments that had agreed to hold dollars but also, and more importantly, because such a move would achieve nothing if the other governments did not revalue against the dollar at the same time. However, as the American reserve position worsened and as inflation and unemployment began to increase, a change of policy leading to a general devaluation became inevitable. On August 15, 1971, President Richard Nixon temporarily froze wages and prices, imposed a 10 percent tariff on imports, and suspended dollar transactions in gold. This was designed to force a devaluation of the dollar on the existing system, while also cooling down the inflationary pressures within the U.S. economy.

The next step was to get the foreign governments to accept a revaluation of their own currencies against the dollar: a general realignment of the parities upward. The idea at this stage was not to create a regime of flexible exchange rates, but to force a general realignment in the parities before they were fixed again. The agreement that finally emerged after several months of hard bargaining at the Smithsonian Institution had to accommodate a 9 percent devaluation of the dollar versus gold and a continued suspension of gold transactions in order to force revaluations in the other currencies. Foreign governments also agreed to undertake a thorough review of the monetary system to achieve a suitable degree of flexibility and a significant liberalization of trade policies that would be acceptable to both sides.

The latter components of the agreement eventually led to the creation of the Committee of Twenty (of the leading OECD economies), which oversaw the attempt to design a system of "stable but adjustable" exchange rates, which was then used to legalize the float that followed under the International Monetary Fund's articles of agreement. The trade liberalization agreement, meanwhile, led to the Tokyo Round of trade negotiations in the later 1970s. The Smithsonian Agreement itself, however, proved to have been too little, too late, as the U.S. trade deficits and reserve position failed to respond. The U.S. economy continued to expand, other countries got into balance of payments difficulties too and floated, and dollars continued to flow into Germany and Japan. In 1973 the United States decided to devalue the dollar again, and one by one the other countries bowed to the inevitable and allowed their currencies to float beyond their new parities. The dramatic increase in energy prices that followed ensured that there was no going back to a regime of fixed parities, although the European Community did manage to create a system of stable but adjustable exchange rates among themselves six years later. In that sense, the Smithsonian Agreement did lead to the flexible exchange rates that we have today and to the European monetary system of 1979 99.

See also Bonn Summit; Bretton Woods system; European Monetary Union; exchange rate regimes; international liquidity; International Monetary Fund (IMF); international policy coordination; international reserves; Plaza Accord

FURTHER READING

Kenen, Peter. 1994. *The International Economy*. 3d ed. Cambridge: Cambridge University Press.

ANDREW HUGHES HALLET

■ smuggling

Smuggling is the surreptitious import or export of goods in violation of domestic or international law. Export smuggling is relatively rare and is mainly relegated to products involved in a multilateral boycott, such as a United Nations embargo on diamonds from African war zones, or state-subsidized products that consequently can be sold for higher prices abroad. Most commonly, smuggling involves the illegal import of banned goods or goods that face import quotas, significant tariffs, or other high taxes.

The preponderance of literature dealing with smuggling in the field of economics focuses on its impact on the welfare of the nation. In the mid-1970s, many believed that smuggling was beneficial, at least to a small country, because smugglers circumvented tariffs that undermined free markets. Consequently smuggled goods (or contraband) could be sold for lower prices and therefore enhanced welfare. Bhagwati and Hansen (1973) challenged this commonly held belief, arguing that the level of prices resulting from smuggling was not the only issue relating to national welfare. Smuggling could be welfare reducing when it undermined government policies of establishing tariffs to protect local industries. Subsequent writers also noted that smuggling could corrupt and demoralize a society as smugglers normally undermined law enforcement through bribery, intimidation, and even murder.

Recent Trends in Smuggling Prior to trade liberalization at the end of the 20th century, most developing countries placed significant tariffs on imports to protect local industries, raise taxes, and/or restrict local demand for scarce supplies of foreign exchange. Although these policies were designed to enhance national welfare, they in fact resulted in higher prices for many products sold in these countries as well as limited product choice. Such policies promoted the growth of widespread smuggling of consumer products into developing countries. Many smugglers proved adept at physically evading law enforcement agents or co-opting them through bribery. Smugglers also found a convenient distribution channel among the many unlicensed street vendors of the developing world's informal economy.

For many in the growing middle classes in these countries, smugglers were seen as benign if not heroic actors. By the mid-1990s, the impact of smuggled goods was substantial in many markets. Contraband in Poland was estimated to equal 20 percent of legally imported goods, and half the computers sold in Brazil were smuggled into the country. In one year alone, Chinese authorities confiscated nearly half a billion dollars in contraband. In developed countries, too, smuggling of certain products thrived. A study by Albers-Miller (1999) observed that consumers of smuggled cigarettes in the United Kingdom considered their actions to be a reasonable response to high cigarette prices.

Widespread trade liberalization in developing countries in the last quarter of the 20th century raised expectations that smuggling would disappear as import quotas were abandoned and tariff levels diminished. However, the relationship of trade liberalization to smuggling proved more complex than originally believed. Gillespie and McBride (1996) studied the impact of trade liberalization on smuggling in Mexico during the 1980s and 1990s. They concluded that 15 years into trade liberalization smuggling had diminished but remained significant.

One reason for the persistence of smuggling after trade liberalization was the continuation of certain cost advantages to smugglers over legitimate importers. Tariffs had been lowered by trade liberalization but still existed, allowing smugglers to profit from circumventing them. For example, in 2006 the Indonesian tire industry was threatened by a surge of smuggled tires from China despite the country's relatively low 15 percent tariff on tires. In addition to tariffs, smugglers could avoid other taxes that their legal counterparts had to pay. In postliberalization Mexico, smugglers who switched to legitimate importation faced paying income tax as well as a 15 percent value-added tax on the products they sold.

In addition, many smugglers protected their low-cost position by increasingly forgoing the smuggling of legitimate goods in order to smuggle stolen or counterfeit goods. Stolen automobiles became especially popular. Cars stolen in the United States were not only smuggled into Mexico but shipped as contraband to China. Counterfeits were another way smugglers lowered the cost of smuggled products by stealing the brand equity of the brand owner. The trend toward smuggling counterfeit goods was exacerbated by greater scrutiny by law enforcement regarding counterfeits crossing borders. Customs authorities of many countries improved their ability to identify and confiscate shipments of counterfeits, thus forcing counterfeiters to seek out smuggling channels.

The diversification of smugglers into other criminal activities signified another disturbing trend. Smuggling was becoming less a business of Robin Hoods who were arguably providing consumers with good products at good prices and more the domain of organized crime. Smuggling activities appeared to be consolidating in the 1990s with contraband gangs involved in moving large shipments of goods. Violence, long associated with organized crime, appeared among these gangs. In Colombia and the United States, smugglers played a new role in the black market peso exchange. After a crackdown on the money laundering of illicit drug profits through U.S. banks, Colombian drug dollars in the United States were used to purchase products that were subsequently smuggled into Colombia. There they were sold for the pesos necessary to support the cartels' production operations. In the late 1990s, Colombian officials estimated that as much as 45 percent of the imported consumer products sold in their country had arrived via the black market peso exchange.

In light of smuggling's failure to disappear with trade liberalization and its increased consolidation and violence, many governments began to attack the problem with more vigor. Penalties, aimed at both smugglers and government employees who accepted bribes from smugglers, were increased. China and Vietnam threatened participants with a newly enacted death penalty. In some countries, such as Mexico, customs authorities were overhauled. Nonetheless, smuggling often remains hard to eradicate due to its organized nature, government corruption, and a lack of resources for law enforcement.

Smuggling and Legitimate Economic Actors
The role in smuggling of legitimate economic actors remains obscure. Legitimate importers and distributors are put at a cost disadvantage when required to compete with contraband. This has led to conflict in the past. In the early 1990s, violent altercations occurred in Mexico City when store owners attempted to evict informal vendors selling contraband outside their stores. However, smugglers and legal distributors need not always constitute two mutually exclusive groups. Several researchers have

proposed that smugglers could camouflage their smuggling with some legal imports and legal distributors could decrease their costs by including some contraband.

There are both benefits and disadvantages for producers of legitimate products that are smuggled. Smuggling works to their advantage when their products are barred from entering markets or when prohibitive taxation discourages purchase of those products. However, multinational corporations can suffer when smugglers undercut legitimate distributors, since legitimate distributors can provide greater oversight of product quality and deliver after-sales service. Relationships with strategic allies, such as licensees or joint venture partners in local markets, can be undermined when contraband enters a partner's market and diminishes legal sales.

The advantages and disadvantages of firms cooperating with smugglers are well illustrated by the case of the cigarette industry. In 1999, internal documents of the five largest tobacco companies were made public as part of a litigation settlement in the United States. These documents revealed a variety of ways in which cigarette producers cooperated with smugglers. These included sending cigarettes to known smugglers, destroying records related to smuggled cigarettes, discussions of how to disguise contraband shipments, setting up international financial accounts to camouflage contraband earnings, and buying air time on transnational media to provide advertising support for smuggled brands.

These revelations spurred a number of national and provincial governments to sue tobacco companies in U.S. courts. Besides seeking compensation for lost taxes and law enforcement costs related to the smuggling, some governments sought compensation for increased health costs due to greater cigarette consumption attributed to the lower price of contraband cigarettes. One suit brought by the European Union (EU) sought compensation of up to $1.7 billion a year. Another suit brought by Colombia charged Philip Morris employees with knowingly participating in the black market peso exchange, using cigarettes to help launder illicit drug money.

Despite the strong evidence of the public record, the plaintiffs were stymied by the U.S. Revenue Rule, which disallowed U.S. courts from assisting foreign governments in their collection of taxes. However, the EU won a ruling that allowed it to present its case anew as one involving money laundering. Subsequent to this decision, Philip Morris agreed to pay the EU more than $1 billion to settle the complaint.

Faced with increased public scrutiny, the multinational tobacco companies began to curtail supplies to distributors involved with smuggling. To the dismay of these multinational companies, smugglers who once helped them sell more of their branded cigarettes in high-tax countries quickly turned to smuggling counterfeit cigarettes instead. Practically overnight, smuggling ceased to be a major contributor to the profits of cigarette companies and became instead a substantial threat to their sales and brands.

Smuggling continues today to undermine public policy and challenge legitimate distribution channels. Trade liberalization has proven to be no panacea. For many governments, the increased association of smuggling with organized crime, counterfeiting, and money laundering only increases the urgency for its eradication.

See also illegal drugs trade; nontariff measures; political economy of trade policy; quotas; tariffs

FURTHER READING

Albers Miller, Nancy D. 1999. "Consumer Misbehavior: Why People Buy Illicit Goods." *Journal of Consumer Marketing* 16 (3): 273 87. Examines why consumers buy contraband and how they justify their behavior.

Bhagwati, Jagdish, and Bent Hansen. 1973. "A Theoretical Analysis of Smuggling." *Quarterly Journal of Economics* 87 (2): 172 87. Presents one of the earliest and most influential discussions of the potential economic impacts of smuggling.

Gillespie, Kate. 2003. "Smuggling and the Global Firm." *Journal of International Management* 9 (3): 317 33. Explores and documents the role of multinational companies in smuggling.

Gillespie, Kate, and J. Brad McBride. 1996. "Smuggling in Emerging Markets: Global Implications." *Columbia Journal of World Business* 31 (4): 39 54. Traces the evolution of smuggling in the face of trade liberalization.

KATE GILLESPIE

■ social policy in open economies

Globalization and other forces are creating increased demand for social policies at the same time as they are restricting the ability of governments to supply such policies. Some observers see in this combination of increased demand and restricted supply a recipe for disaster in which globalization increasingly undermines the ability of governments to provide for the social well-being of their citizens. Both the theory and evidence available to date, however, suggest the glass is more half-full than half-empty: effective social policies remain alive and well in open economies, and the positive effects of the restrictions on governments' ability to provide social policy outweigh the negative effects.

The purpose of this entry is to outline the forces that increase the demand for social policies, the forces that restrict the ability of governments to supply such policies, and the conditions under which social policies can be provided. The concept of social policy as used here refers to the range of government policies that provide a social safety net, including employment standards involving factors such as minimum wages, hours of work, and terminations; health and safety regulations and programs, including workers' compensation, human rights and antidiscrimination legislation, and transfers such as unemployment insurance and welfare. Particular attention is paid to the United States and Canada, given the greater openness that has occurred in those countries after a number of recent free trade agreements.

Forces Increasing the Demand for Social Policy Initiatives The demand for social policy initiatives is increasing in part because of the interaction of three sets of interrelated forces: those affecting the demand side of the economy, especially the labor market; those affecting the supply side, especially related to demographics; and those affecting institutional changes.

On the demand side, labor markets are affected by a variety of interrelated forces: globalization and freer trade; technological change, the computer revolution, and the shift to a knowledge-based economy; industrial restructuring, especially from manufacturing to services; and deregulation and privatization. These pressures in turn give rise to a variety of manifestations, many of which increase the demand for social policy initiatives. Wage inequality has increased (Cline 1997; Rodrik 1997) as the lower end of the wage distribution has been adversely affected by various factors including import competition from low-wage countries, skill-biased technological change, outsourcing, and the supply influx from displaced workers who have lost their former blue-collar, male-dominated, often unionized jobs in manufacturing. Nonstandard, or "contingent," employment has increased in various forms (part-time, limited-term contracts, temporary-help agencies, self-employment, home work, and telecommuting). Many of these jobs have effectively shifted risk from employers to employees and created groups of vulnerable workers often without pensions or health care.

Numerous factors on the supply side are also increasing the demand for social policy initiatives. The population and labor force are aging and living longer, with the large baby-boom cohort entering the ages of retirement where disability and health care costs increase dramatically. The aging "gray panthers" are creating a voting constituency for social policies related to pensions, social security, and age discrimination policies that often have to be financed by a smaller cohort of taxpayers in younger generations. Many are facing issues of elder care for their own ailing and longer-lived parents. The rise of the two-earner family means an increased demand for social policies in areas of institutional elder care and childcare as well as for "family friendly" policies in areas related to working time and parental leaves. The rise of single-parent families has increased the demand for state support as a surrogate parent. The growing diversity of the workforce is creating demands for policies in the areas of human rights, antidiscrimination, and reasonable accommodation for disabled persons.

At the same time as these demand- and supply-side forces are increasing the demand for social policy initiatives, various institutional changes are also working in the same direction. The decline of unionization is creating a vacuum in areas where unions once provided protection and a safety net for unionized workers. The declining real value of the minimum wage has left many lower-wage workers increasingly vulnerable (DiNardo, Fortin, and Lemieux 1996). The decline of employer-sponsored occupational pension plans has exposed workers to more insecurity with respect to their retirement income, and the shift from defined-benefit to defined-contribution pension plans has exposed them to more investment risk.

While these various forces have often increased the demand for social policy initiatives, in other ways they have also reduced that demand. The two-earner family with women increasingly participating in the labor force obviously provides greater income potential and insurance against the risk of income loss associated with one member being unemployed or underemployed or disabled. Nonstandard employment can provide the flexibility for families to provide childcare or eldercare. The ability of older workers to continue working (in part due to legislative bans on mandatory retirement) can provide the potential for retirement income. The application of human rights and antidiscrimination policies as well as employment standards can facilitate individuals' earning income and thereby reduce their reliance on social policies provided by governments.

Forces Inhibiting Governments from Supplying Social Policies Many of the same forces that are increasing the demand for social policies are also reducing the ability of governments to finance social policies. Specifically, under globalization, freer trade, and falling transportation and communication costs, companies are more able to relocate their plants and investment to low-cost jurisdictions and export their goods and services back into higher-cost jurisdictions. Regulatory costs, including those arising from social policies, can be an element of those cost considerations. Such costs can arise from various regulations: labor standards (e.g., such areas as minimum

wages, hours of work and overtime, vacations, holidays and leave, terminations, severance pay, and unjust dismissal protection); collective bargaining regulations; antidiscrimination; reasonable accommodation requirements; workers' compensation; and health and safety regulations.

Because of the threat of the loss of business investment and the jobs associated with that investment, governments are under increasing pressure to compete in part by reducing their regulatory costs, including those arising from social policies. In essence, they are under more pressure to be "open for business." This is the case for governments across different countries as well as for local governments within countries.

Phrases used to describe this phenomenon generally have negative connotations: *social dumping, the rule of the market over the rule of law, harmonization to the lowest common denominator, regulatory meltdown, race to the bottom*, and *ruinous competition.*

Forces for Downward Harmonization to Occur
For such harmonization to occur and for it be to the lowest common denominator (i.e., to the jurisdiction with the least costly regulations and social policies) a number of conditions have to prevail (Gunderson 1998). First, the regulations and policies have to be enforced; otherwise there is no effective cost on employers. In many cases, initiatives that appear extremely costly "on the books" are simply not extensively enforced in practice. In this case, the intended benefit is also not realized.

Second, for the initiatives to be costly, the benefits to employers must be less than the costs. In many cases, employers themselves benefit from the initiatives, and this can offset at least part of the costs. Workers' compensation, for example, imposes a payroll tax on employers, but it also frees them from the threat of being sued by injured workers since workers gave up that right as the quid pro quo for essentially receiving "no-fault" coverage in the event of a workplace injury. Policies that require employers to give advance notice in the case of massive layoffs or plant closings can be costly, but they can also facilitate job searches that can benefit other employers.

Third, employers may shift much of the cost of work-related policies forward to customers or backward to workers. Shifting costs forward to customers is increasingly difficult in a world of global competition since customers can purchase from lower-cost sources that do not have such extensive regulatory costs. Shifting costs backward to labor, however, is possible since labor tends to be the immobile factor of production and hence cannot easily "escape" the cost shifting by moving to employers that do not shift their regulatory costs. Such cost shifting can occur in the form of lower compensating wages paid in return for the benefit workers receive from the policy. In other words, workers who receive those benefits effectively pay for them through receiving lower wages in return for the benefits. Evidence suggests, for example, that the majority of payroll taxes initially levied on employers for policies such as workers' compensation, pensions, and unemployment insurance are ultimately shifted back to workers in return for the expected benefits they receive (Kesselman 1996). In that vein, some of the populist views of imposing regulatory costs on large wealthy corporations are somewhat misguided in that the vast majority of those costs are shifted backward to workers in return for the benefits they receive from the regulations. There is no such thing as a "free lunch" in this area.

Fourth, for the harmonization to be downward, employers must respond to the cost increases by threatening to move their plants and investments to other, lower-cost jurisdictions. Although such a threat is now more credible, given the global conditions outlined previously, business investment and plant location decisions respond to a myriad of factors other than the cost of social programs. Furthermore, social programs can provide benefits to employers in forms such as infrastructure and political and social stability (Rodrik 1997).

Fifth, for downward harmonization to occur, governments must respond to the threat of capital mobility by reducing their costly social policy initiatives. Although there will be pressure in this direction, some governments (reflecting the prefer-

ences of their constituencies) may simply be willing to "pay the price" of their regulation. They may opt to be a "kinder and gentler society" and sustain their social safety net even if it costs them business investment and the associated jobs. They may hope to retain and attract what they regard as socially responsible business.

While these links must all be present for interjurisdictional competition for investment and the associated jobs to lead to downward harmonization to the lowest common denominator, the fact remains that the pressures are in the direction of such harmonization. That is, to a large degree each of the links is present: most policy initiatives are enforced; they impose net costs on employers; employers are not able to fully shift the costs; they have a credible threat to move to other jurisdictions that do not impose such costs; and governments do respond to the need to attract investment and jobs. As such, the pressures are in the *direction* of downward harmonization. The question is: How much? And are there countervailing pressures, often associated with openness and globalization, working against downward harmonization?

Forces for Upward Harmonization Although various forces foster downward harmonization, other forces, often associated with globalization and open economies, foster upward policy harmonization.

Consumer groups and nongovernmental organizations, often internationally organized and coordinated through the Internet, can put pressure on brand-name multinationals through various mechanisms including consumer boycotts, "Internet outings," and requiring a "social label" indicating the conditions under which the product was produced.

Multinationals themselves can foster upward harmonization by "exporting" the more advanced practices they follow in their home country and applying a more uniform set of their company practices across their different operations. Multinationals associated with brand names are especially sensitive to their public image and want to appear as good corporate citizens in the countries in which they operate. They often operate under voluntary corporate codes

of conduct and follow International Labor Organization/Organisation for Economic Co-operation and Development guidelines.

Open economies can also foster upward harmonization of social policies through the use of social clauses and side accords in trade agreements, such as the labor and environmental side accords under the North American Free Trade Agreement and the Social Charter of the European Union (EU). In such circumstances, countries that may not have enforced their labor standards are required to do so or even to raise them to the higher standards of trading partners, perhaps with the aid of funds such as the EU's Social Funds.

Social policies may also be harmonized upward through the emulation of best practices in the policy arena. To the extent that openness and globalization foster income convergence, the more rapid growth of the poorer countries may also enable them to afford more extensive social policies, again fostering upward harmonization. Empirical evidence supports the proposition that as income increases, the demand for social policies also increases (e.g., Fields 1995). Rodrik (1997) also emphasizes how social policies that assist those who are adversely affected by market-oriented policies such as free trade can foster efficiency by reducing public resistance to such market-oriented policies; equity and efficiency need not always conflict.

Forces for Sustained Divergence Various forces are also at work to sustain a divergence in social policies. Lower-income countries, for example, may be caught in a "low equilibrium" trap rather than converging upward. A divergence of social policies may therefore be sustained for an extended period of time, just as divergent growth rates may be sustained.

The conditions under which social policies are first established may exert a persistent influence on future policies. This is especially the case when bureaucracies and interest groups have a vested self-interest in sustaining the policies, and they become embedded in social structures and cultural norms within borders. Seymour Martin Lipset, for example, argued that the United States was initially influenced

by the American Revolution, which fostered a path of individualism and distrust of government. Canada, in contrast, started off on a path of loyalty to Britain and less distrust of collective action through the state. Such different initial conditions and subsequent path dependence may help explain the more extensive social policies that prevail in Canada compared with the United States.

Sustained divergence can also be fostered by the heterogeneous demands of various groups with differences in their preferences or needs for social policies as well as in their willingness and ability to pay for such policies. In such circumstances, a Tiebout-type equilibrium may prevail where individuals and firms sort themselves into jurisdictions that offer the social expenditure tax combination that meets their preferences. Just as openness and globalization can foster diversity in consumer options, they can foster diversity in choices of social policies.

Evidence: Harmonization or Sustained Divergence Clearly, theoretical reasoning suggests that globalization and openness along with other forces can foster harmonization in any direction: downward to the lowest common denominator, upward to a higher standard, or sustained divergence. As is commonly the case, theory covers all bases; hence the need for an appeal to the evidence.

The evidence, however, is not clear-cut. Gomez and Gunderson (2005) review the international evidence and particularly that of the United States and Canada. They conclude that increased openness has fostered a *tendency* toward harmonization of social policies and that harmonization has been *toward* the lower common denominator. Between the United States and Canada, for example, there has been recent downward harmonization toward the lower U.S. level in areas such as collective bargaining laws, minimum wages, unemployment insurance, equal pay laws, pension funding, and welfare and family benefits for working families. There has been sustained divergence in workers' compensation and health and safety, however, and some upward harmonization to the higher U.S. standards with respect to age discrimination. They also cite a wide range of studies, especially by political scientists (e.g., Garnett

1998), who tend to find more sustained divergence and emphasize that governments still maintain considerable degrees of freedom in setting their social policies in spite of greater openness and competitive pressures.

As indicated, Fields (1995 and other studies cited therein) have examined poor countries and indicated that as their incomes grow their demand for social policies also increases; that is, social policies are "normal goods" that are collectively desired and that can be afforded as income grows. In essence, the upward convergence of the income of poor countries toward that of wealthier countries also fosters an upward convergence in their social policies. It follows that prematurely imposing costly regulations on developing countries for example, as a condition of a free trade agreement can dampen the very growth process that would foster a natural upward convergence of social policies.

Countries of the EU did not experience a downward harmonization in their social policies as a result of the greater trade and capital and labor mobility after the formation of the union. This reflects the fact, however, that the EU Social Charter required the poorer countries such as Greece and Portugal to harmonize upward as a condition of entry into the EU. The EU also provided Social Funds to the poorer countries to facilitate their upward harmonization (Gomez and Gunderson 2005).

Even if there is a tendency toward downward harmonization, this should not automatically be regarded negatively. With greater openness, governments are simply under more pressure to pay attention to the cost consequences of their social policies; they face a "harder" rather than "softer" budget constraint. In such circumstances, social policies that are inefficient and involve mainly rent seeking (e.g., possibly procurement policies) are under the most pressure to dissipate since they do not foster the competitiveness that enables countries to ultimately afford their social policies. The competitive position of employers can be jeopardized by such policies, and their costs cannot be shifted back to workers since they do not have associated benefits for workers. In contrast, social policies that have positive feedback

effects on efficiency and support a public infrastructure should survive indeed thrive under greater openness and competitive pressures since they foster competitiveness. The area of greatest concern, however, is whether governments can sustain pure equity-oriented social policies that do not have positive feedback effects on efficiency but protect the most vulnerable and disadvantaged groups in society. This is the main challenge to social policy under increased openness and competitiveness.

See also globalization; labor standards; trade and the environment

FURTHER READING

Cline, W. 1997. *Trade and Wage Inequality*. Washington, DC: Institute for International Economics. Discusses the theory and evidence behind how trade, globalization, skill biased technological change, and other factors have led to growing wage inequality.

DiNardo, J., N. Fortin, and T. Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973 92: A Semi Parametric Approach." *Econometrica* 64: 1001 44. Discusses the role of changing labor market institutions such as declining unionization and declines in the real value of minimum wages in growing wage inequality.

Fields, G. 1995. *Trade and Labor Standards: A Review of the Issues*. Paris: OECD. Outlines the effect of labor standards on trade as well as how trade fosters growth in developing countries and how this in turn fosters higher labor standards.

Garnett, G. 1998. *Partisan Politics in a Global Economy*. Cambridge: Cambridge University Press. Provides a political science view of how governments have been able to maintain their role in the face of rising of global market pressures.

Gomez, Rafael, and Morley Gunderson. 2005. "Does Economic Integration Lead to Social Policy Convergence? An Analysis of North American Linkages and Social Policy." In *Social and Labor Market Aspects of North American Linkages*, edited by Rick Harris and Thomas Lemieux. Calgary: University of Calgary Press, 309 56. Discusses the empirical evidence on the harmonization of social policy, especially between Canada and the United States.

Gunderson, Morley. 1998. "Harmonization of Labor Policies under Trade Liberalization." *Relations Industrielles/ Industrial Relations* 53 (1): 11 41. Outlines the pressures for the harmonization of labor laws and policies as well as the conditions under which there will be downward or upward harmonization.

Kesselman, J. 1996. "Payroll Taxes in the Finance of Social Security." *Canadian Public Policy* 22 (June): 162 79. Outlines the theory and evidence with respect to the cost shifting of payroll taxes used to fund social programs.

Rodrik, D. 1997. *Has Globalization Gone Too Far?* Washington, DC: Institute for International Economics. Emphasises the important role of the state in fostering labor and social standards that in turn reduces the resistance to globalization.

MORLEY GUNDERSON

■ **source country**

See multinational enterprises

■ **South-South trade**

South-South trade (SST) refers to trade between developing countries, which represent a limited share of the world economy in terms of supply capacity and markets. Trade between developing countries has been steadily limited despite the supposed gains associated with better matching of supply and demand characteristics and easier access due to less stringent standards, which should promote regional integration. Contrary to the common perception, however, in recent years SST has been a major contributor to the growth of world trade. Trade between developing countries accounted for 18 percent of world trade in 2005, an increase of 5 percentage points within five years. In contrast, the share of North-North shipments in world trade has been reduced by 5 percentage points over the same period.

China is the most prominent country in this regard. On the import side, China has become in 2005 the second largest market for exports from the South behind the United States, but ahead of Japan or

Germany. Moreover the pace of growth is impressive, with a 232 percent increase of Chinese imports from other developing countries within five years, excluding oil. Besides China, a series of developing countries contribute to SST by offering buoyant markets: essentially Asia (Hong Kong, Singapore, Korea, Taiwan, Malaysia, Thailand, India, etc.) and Mexico.

Still, a large number of developing countries have been missing this general trend, in particular the least-developed countries, and a key pattern of the recent increase is that it has been very unevenly distributed among countries in the South. The sectoral pattern of exports very much reproduces the general pattern of specialization of the developing countries. It does not come as a surprise, therefore, to see China importing natural resources from other developing countries.

Explanations of a deficit of South-South trade (SST) range from geography, institutions, and trade costs to the difficulty of pursuing regional integration in the South. The asymmetric nature of trade liberalization may have played a role too: preferences have been conceded by the North on a nonreciprocal basis, without requesting access to markets in the South. This approach has made it more appealing for the South to trade with the North than with the South (Bouët, Fontagné, and Jean 2006).

Vertical Division of Labor and Extensive Margin of Trade Two main determinants of the recent dynamism of SST must be stressed. First, the vertical specialization of countries (Hummels, Ishii, and Yi 2001) and the associated increasing fragmentation of the production processes translate into large flows of parts and components among developing economies. Vertical specialization accounts for roughly one-third of world trade growth. Such processing trade is well documented in the case of Asia. China, especially, is often used as an export platform, importing and assembling intermediate goods produced by affiliates of foreign firms located elsewhere in Asia.

Second, the diversification of numerous developing economies has led them to compete on a wide spectrum of products with industrialized countries. According to the distinction introduced by Melitz

(2003), the recent dynamism has been based on the extensive margin of developing countries' exports (that is, increasing the number of exported products or destination markets), rather than on the intensive one (exporting increasing values of the same products to the same markets).

Using detailed U.S. import data, Hummels and Klenow (2005) construct measures such that a country's share in world exports decomposes in an extensive export margin (the fraction of world exports occurring in the product-market categories where a country exports), times an intensive export margin (a country's share of world markets in the market-categories in which it exports). Though not specifically addressing SST, such a decomposition is very informative. The underlying question is: As a country develops and accumulates resources, how are these resources used? To increase the quantities exported of the existing set of goods? To improve the quality of these goods? Or to expand the set of goods or destination markets? The bottom line is that two-thirds of the growth pertains to the extensive margin, essentially due to a larger number of exported products: A simple count points to a 62 percent increase in their number when the economic size of the exporter doubles.

In 2005 China overtook France and the United States in terms of diversity of exports. Other highly diversified exporters include Korea, Taiwan, India, Turkey, Thailand, Brazil, Mexico, Malaysia, Singapore, Hong Kong, Indonesia, South Africa, United Arab Emirates, Argentina, and the Philippines. In contrast, Ethiopia has simply kept its number of products constant over the period 2000 2005, while Jamaica, Myanmar, El Salvador, Mali, and Côte d'Ivoire have recorded declines in the number of their exported products. Still, the question remains whether recorded SST flows remain too low, given the characteristics of these countries.

The Gravity Puzzle Using gravity equations to assess deviations between actual and predicted trade has been the most common approach to study SST. The augmented gravity equation explains the value of bilateral trade by the gross domestic products (GDPs) of exporter and importer, the distance be-

tween them, and a vector of control variables accounting for the level of trade resistance between them, such as the level of tariffs, whether they share a common language or a common border, and the existence of a free trade agreement (FTA). In the case of SST, the appealing feature of this approach is that the data involved are readily available.

This approach may be flawed, however, due to omitted variables and endogeneity problems, as examined in Baier and Bergstrand (2006). The latter type of problem is particularly penalizing, since it sheds doubt on the extensive literature addressing the trade impact of integration among developing economies. In a nutshell, are FTAs randomly formed, or are they formed among countries that share characteristics that make them good candidates for integration? If the latter, FTAs' impact on trade might well be overestimated.

Examining intra-WAEMU (West-Africa Economic and Monetary Union) trade, Coulibaly and Fontagné (2006) stress the combined role of economics and geography in the persistence of untapped trade potentials. Being poor and landlocked profoundly reduces trade: beyond distance, the worse combination for SST is accordingly to export from a landlocked country, through a transit country, with a limited share of paved roads.

Poor infrastructure is, however, only a partial explanation for the low level of SST, since it also affects domestic trade. Low levels of SST also result from tariffs, nontariff barriers (NTBs), preferences, and more generally all aspects that make it more difficult to trade with another country than with another region within the same country. Capturing such effects imposes a methodological shift: the right benchmark is domestic trade (the difference between domestic production and exports). A growing literature is using the trade and production database of the United Nations Industrial Development Organization and the World Bank to analyze sectoral trade and production data for a series of countries in the North and in the South. This makes it possible to use the so-called border effect methodology to assess whether something specific is hampering SST.

Controlling for supply capacity, demand, distance, tariffs, NTBs, and common colonial power, SST is much more deflected by the existence of borders than North-North trade or North-South trade are. Mayer and Zignago (2005) estimate that the impact on SST of the barriers (custom formalities and delays, differences in preferences or regulations, etc.) associated with the existence of a border is the equivalent of a 100 percent tax on imports, and this effect is in addition to that of any existing tariffs and NTBs.

In this context, do free trade areas offer a viable solution? Fontagné and Zignago (2007) address this issue in a study designed to control for systematic differences in trade across countries and productive sectors. The six most prominent preferential trade agreements (PTAs) according to their trade impact are the Central American Common Market, the Andean Community, the North American Free Trade Agreement, the European Union, Mercosur (the regional trade agreement among Brazil, Argentina, Paraguay, and Uruguay), and the Association of Southeast Asian Nations; other PTAs have a smaller impact on trade.

Another route to trade liberalization among developing countries proceeds from their active participation in the multilateral arena. Simulations conducted with computable general equilibrium models conclude that this is key to the "development part" of the ongoing multilateral round of negotiations (Francois, van Meijl, and van Tongeren 2005).

Recent literature reexamines the common view that trade promotes peace and finds that the impact of bilateral and multilateral trade is not the same. Considering military conflicts in the period between 1950 and 2000, the probability of a conflict is lower among countries trading more bilaterally, because of the larger trade losses incurred in case of conflict (Martin, Mayer, and Thoenig 2007). Hence regional agreements in the South, as opposed to the multilateral trade liberalization agreements, reinforce the bilateral dependence among developing countries and reduce the probability of occurrence of a conflict.

See also gravity models; Mercosur; North-South trade; trade and economic development, international

FURTHER READING

Anderson, James, and Eric van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review* 93 (1): 170 92. The seminal paper regarding the right formulation of the augmented gravity equation.

Baier, Scott L., and Jeffrey Bergstrand. 2006. "Do Free Trade Agreements Actually Increase Members' International Trade?" *Journal of International Economics* 71 (1): 72 95. Correctly addresses endogeneity issues in assessing the trade impact of PTAS.

Bouët, Antoine, Lionel Fontagné, and Sébastien Jean. 2006. "Is Erosion of Preferences a Serious Concern?" In *Agricultural Trade Reform and the Doha Development Agenda*, edited by Kim Anderson and Will Martin. Washington, DC: Oxford University Press and The World Bank, 161 92. First exhaustive examination of the impact of conceded preferences, using detailed data.

Cheptea, Angela, Guillaume Gaulier, and Soledad Zignago. 2005. "World Trade Competitiveness: A Disaggregated View by Shift Share Analysis." CEPII Working Paper 2005 23. World trade at a glance for some 5,000 products.

Coulibaly, Souleymane, and Lionel Fontagné. 2006. "South South Trade: Geography Matters." *Journal of African Economies* 15 (2): 313 41. Derives and estimates a model highlighting the impact of geography and infrastructures on bilateral trade flows within the West African Economic and Monetary Union.

Feenstra, Robert C. 1998. "Integration of Trade and Disintegration of Production in the Global Economy." *Journal of Economic Perspectives* 12 (4): 31 50. How the value added chain is decomposed internationally.

Fontagné, Lionel, and Soledad Zignago. 2007. "A Re Evaluation of the Impact of Regional Agreements on Trade Patterns." *Économie internationale* 109: 31 51. Estimates the impact of regional agreements using a panel of 135 countries trading in 26 industries over the period 1976 2000.

Francois, Joseph, Hans van Meijl, and Frank van Tongeren. 2005. "Trade Liberalization in the Doha Development Round." *Economic Policy* 20 (42): 349 91. The round from a development perspective at a glance.

Hummels, David, Jun Ishii, and Kei Mu Yi. 2001. "The Nature and Growth of Vertical Specialization in World Trade." *Journal of International Economics* 54 (1): 75 96. Quantifying the role of trade in intermediate goods in the observed world trade growth.

Hummels, David, and Peter J. Klenow. 2005. "The Variety and Quality of a Nation's Exports." *American Economic Review* 95 (3): 704 23. Do countries allocate their new resources to increase the quantity of exports of existing products or to enlarge the portfolio of their exported products?

Martin, Philippe, Thierry Mayer, and Mathias Thoenig. 2007. "Make Trade not War?" Mimeo. Paris School of Economics. Available at http://team.univ paris1.fr/teamperso/martinp/war april2007.pdf. New evidence explaining why globalization has not lived up to its promise of decreasing conflict.

Mayer, Thierry, and Soledad Zignago. 2005. "Market Access in Global and Regional Trade." CEPII Working Paper 2005 02. Border effects, the direction of trade flows, and the trade impact of PTAs.

Melitz, Marc J. 2003. "The Impact of Trade on Intra Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695 725. The seminal paper distinguishing between intensive and extensive margins of trade.

LIONEL FONTAGNÉ

■ sovereign risk

Sovereign risk refers to circumstances in which governments default on loan contracts with foreigners, seize foreign assets located within their borders, or otherwise prevent domestic holders of foreign capital from meeting obligations. Because there is no supranational authority that can enforce contracts across borders, when sovereigns choose not to honor contracts with foreign investors, those investors have little recourse to recoup their losses. Consequently relationships between foreign investors and sovereigns are dictated primarily by the sovereign's willingness to pay rather than its ability to pay.

Foreign governments could become better credit risks by waiving sovereign immunity. The legal doctrine of sovereign immunity can be interpreted as exempting the property of foreign governments

from the jurisdiction of domestic courts. Historically, this doctrine has sometimes constrained the ability of creditors to sanction foreign governments that have defaulted. The doctrine has evolved over time to grant creditors some recourse against defaulting sovereigns, however.

History is replete with instances of sovereign default. During the 19th and 20th centuries, in many instances the repayments of loans to sovereigns were far from what was called for in contractual obligations. In the 1820s, for example, Latin America experienced a wave of defaults among its newly independent countries. Many Latin American countries defaulted again in the 1870s, as did Egypt and Turkey. In the 1930s, most sovereign debtors suspended interest payments in the wake of the global depression.

Sanctions and Reputation The risk that a sovereign will not honor its obligations with foreign investors is mitigated by two factors: the threat of sanction and the loss of reputation. Direct sanctions may involve significant costs for defaulting countries. For example, sanctions could take the form of intercepting payments that the sovereign might make to exporters or payments it might receive from importers. Potential exporters to the country might then be less willing to supply the debtor country, knowing that their compensation might be intercepted. Generally, however, the net gain to creditors from sanctions will be less than the cost to the debtor. In theoretical models of sovereign borrowing with sanctions, the optimal amount of borrowing is constrained by the sanction cost: the parties to the debt contract write an incentive-compatible contract that never calls on the sovereign to make a payment to foreign creditors in excess of the sanction cost. This incentive-compatibility constraint leads to less lending than would be the case in a world where no country ever defaulted. An implication of this line of theorizing is that stronger, credible sanctions would allow sovereigns to increase the amount of their borrowing.

Reputation considerations provide another motive for sovereigns to honor obligations with foreign investors. A sovereign might choose to repay in order to maintain continued access to international capital markets on favorable terms. This idea was articulated in work that showed formally the conditions under which countries repay debts in order to preserve a reputation for repayment (Eaton and Gersovitz 1981). Subsequent work questioned whether reputational considerations alone were enough to sustain international lending in the absence of effective sanctions (Bulow and Rogoff 1989). If defaulting countries are able to earn a market return on their savings abroad they could, in theory, borrow, invest the funds abroad, default on the loan, and use the proceeds of their investments abroad to finance domestic investment opportunities in perpetuity.

Reputation models of international lending were revived by Cole and Kehoe (1997), who countered that reputation alone may support international lending if reputational spillovers are important enough. They argue that sovereigns participate in many different relationships. If they renege in one of those relationships, the damage may have adverse consequences for other relationships. If the spillovers are large enough and last long enough, then reputation considerations alone cannot support international borrowing and lending, without resort to the threat of direct sanctions. In general, though, if countries with poor credit histories are able to borrow abroad without fear of their assets being confiscated, the threat of reputation loss becomes less of a factor in deterring default.

Causes and Consequences of Default Risk Several factors have been found to influence the likelihood of sovereign default. Historical data suggest that the risk premium on foreign debt (over the return on U.S. Treasury bonds or British consols) increases with the size of a country's trade deficit and budget deficit. In addition, internal political considerations play a role: those countries that respond to a crisis by raising taxes or cutting spending are less likely to default. Severe external shocks have been found to precipitate default (Diaz-Alejandro 1983). In addition, the debt-to-export ratio and deteriorating terms of trade appear to influence the likelihood of sovereign default. In general, a country's internal

and external circumstances play a role in sovereign risk.

The tangible consequences of sovereign default have proven difficult to uncover. The economists Lindert and Morton (1989) examined data on the 10 leading debtor countries from 1850 to 1970 and found that, on average, foreign lending earned a higher return than domestic lending in the United States and the United Kingdom. Eichengreen and Portes (1989) examined data on bond issues in the 1920s and found that for bonds issued in London, foreign lending earned a higher return than on contemporaneous British consols. For loans originating in the United States, overseas lending earned a return only slightly lower than that on U.S. Treasury securities. Actual yields on foreign lending compensated British and U.S. investors for suspension of payments and write-downs of principal.

When countries default, have they been denied access to international capital markets? Here the evidence is somewhat mixed. Countries that did not default in the 1930s appear not have gained wider access to international capital in the 1930s, or following World War I. Indeed, default of some countries led to a shrinking of global capital flows to all, as the international capital market virtually shut down. A review of historical evidence suggests that there is little, if any, evidence to suggest that the volume of foreign funds a sovereign could borrow was adversely affected by a record of prior default (Eichengreen 1991).

Why do lenders keep lending to countries that have a history of default? Lenders are often uncertain about the characteristics of borrowers. Consequently, new information about borrowers can have a large impact on lending. In the case of countries, a credible change in policy regime may negate the influence of a history of default. If a country changes its regime to focus on economic stability and sound monetary and fiscal policies, history suggests that lenders will be willing to extend further credit.

See also asymmetric information; contagion; currency crisis; international financial architecture; Latin American debt crisis; original sin; peso problem; spillovers

FURTHER READING

Bulow, J., and K. Rogoff. 1989. "Sovereign Debt: Is to Forgive to Forget?" *American Economic Review* 79: 43 50. Shows how lending must be supported by direct sanctions available to creditors, and not by reputation for repayment.

Cole, H., and P. Kehoe. 1997. "Reviving Reputation Models of International Debt." *Federal Reserve Bank of Minneapolis Quarterly Review* 21: 21 30. If a country is involved in relationships with creditors that provide enduring benefits, reputation alone may be enough to support lending arrangements.

Diaz Alejandro, C. 1983. "Stories of the 1930s for the 1980s." In *Financial Policies and the World Capital Market*, edited by Pedro Aspe Armella, Rudiger Dornbusch, and Maurice Obstfeld. Chicago: University of Chicago Press, 5 40.

Eaton, J., and M. Gersovitz. 1981. "Debt with Potential Repudiation: Theoretical and Empirical Analysis." *Review of Economic Studies* 48: 289 309. Analyzes borrowing in the absence of explicit penalties for nonpayment.

Eichengreen, B. 1991. "Historical Research on International Lending and Debt." *Journal of Economic Perspectives* 2: 149 69.

Eichengreen, B., and R. Portes. 1989. "After the Deluge: Default, Negotiation, and Readjustment during the Interwar Years." In *The International Debt Crisis in Historical Perspective*, edited by B. Eichengreen, and P. Lindert. Cambridge: MIT Press, 12 37.

Lindert, P., and P. Morton. 1989. "How Sovereign Debt Has Worked." In *Developing Country Debt and Economic Performance*, vol. 1, *The International Financial System*, edited by J. Sachs. Chicago: University of Chicago Press, 39 106. Historical analysis of the effect of sovereign default on interest rates.

Ozler, S. 1993. "Have Commercial Banks Ignored History?" *American Economic Review* 83: 608 20. Investigates how past defaults affected loan terms for developing countries during the 1970s. The paper finds evidence that lenders did take default history into account when issuing loans in the 1970s. Countries with a poor record of debt repayment faced higher commercial bank lending rates in the 1970s.

KEITH SILL

■ sovereign wealth funds

Many capital-exporting developing countries look for systematic ways of raising returns on their international currency reserves on a long-term basis by creating sovereign wealth funds (SWFs), which are designated pools of assets owned and managed by governments and predominantly deployed worldwide to attain higher returns.

Countries usually have their international reserves managed by their central banks and held in liquid assets in reserve currencies. Since the primary functions of international reserves are to finance payment imbalances and limit exchange rate volatility, the reserves must have a high degree of liquidity. Since assets typically have lower rates of return the more liquid they are, however, it can make sense for governments to invest in longer term, less liquid assets such as bonds and equities. Similarly there may be gains from diversification by investing in a broader range of countries than only those with major reserve currencies.

Perhaps the clearest rationale for the establishment of an SWF is to accumulate reserves in countries that are major exporters of nonrenewable resources such as oil and gas. Concern for future generations dictates that all current revenues from the sale of such commodities not be spent on current consumption. Nor, typically, could all of the remainder be spent productively on domestic investment. Usually a substantial proportion of these revenues accrue to governments from tax receipts or direct ownership. Thus governments accumulate reserves.

Oil-producing countries make up more than half of all SWF funds (i.e., commodity-based funds) in terms of assets under management. Kuwait's International Authority, funded by oil, was established in 1953 and is the oldest SWF. Another prime example is Norway's Global Pension Fund, funded by a portion of North Sea oil and gas. The United Arab Emirates' Abu Dhabi Investment Authority Fund, which was established in 1976, is the world's largest SWF currently, with U.S. $625 billion under management. Other oil exporters, Oman and Brunei, also created investment agencies to recycle their reserve holdings in the 1970s and 1980s, and Russia followed suit more recently, creating a Stabilization Fund in 2003 (Lyons 2007).

Among the better-known non-commodity-based SWFs in Asia are Singapore's Government Investment Corporation (GIC) and Temasek Holdings. Since some of the funding sources for the agencies also include pension contributions from Singapore residents, however, these entities are strictly speaking a combination of SWFs and sovereign provident funds. GIC, established in 1981, has around U.S. $215 billion under management and tends predominantly to make financial and real estate investments. Temasek, established in 1974, has around U.S. $100 billion under management and is a more active investor in international companies regionally and globally.

Although SWFs have been around since the 1950s, they have only recently attracted much public attention. SWFs have taken on increased prominence with their phenomenal growth in recent times (both in numbers as well as funds under management), especially with the creation of the China Investment Corporation (CIC) in 2007. The CIC, which is said to be modeled on Singapore's GIC in both concept and design, began operations when the Chinese government transferred U.S. $200 billion of its U.S. $1.3 trillion in reserves to the agency, making it the world's fifth largest SWF. The new fund's first major investment was a $3 billion investment in the U.S.-based Blackstone private equity group. SWFs took center stage at the October 2007 World Bank and the International Monetary Fund (IMF) annual meetings, and the G7 industrial countries have begun to pay greater attention to these entities.

Since many of the SWFs practice considerable secrecy, the total value of their assets is not known, but estimates put the figure at between $2 and $3 trillion, substantially larger than the global aggregate of all hedge funds. Estimates also project a rapid rate of growth, with figures of $10 trillion and more expected to be reached by early in the 2010s.

Another rationale for SWFs is to minimize the destabilizing effects of fluctuations in exports due to periods of more and less rapid growth in the world

economy. Medium-term fluctuations are common for a wide range of agricultural as well as both renewable and nonrenewable raw materials, so economists have long recommended that governments should accumulate international reserves in good times and draw them down in bad. This has proven difficult to do in many developing countries because of political pressures to spend all increases in reserves. These tendencies can be partially countered through the establishment of SWFs. An example is Chile's Economic and Social Fund, established in 2006.

A third rationale for the creation of SWFs is to eliminate a continual surplus. When continuing disequilibrium in a country's balance of payments has led it to accumulate far more international reserves than it needs based on traditional criteria, and the country expects to continue to accumulate reserves (or at least, not have huge reductions), then it makes no sense to hold all of its reserves in low-yield liquid assets. The country could put its excess reserves into a new facility to earn higher returns, but if reserves are clearly excessive then it would most likely take actions to eliminate its continual balance of payments surplus. Only if the rate of return on domestic investment was quite low would it make sense for the government to continue to accumulate international financial assets, even if these were invested in less liquid, higher return securities. In a market-oriented economy without capital controls, private investment would tend to flow from lower- to higher-yield areas. Thus, if domestic investments were indeed lower yield, capital would flow out and eliminate the balance of payments surplus.

A prime example of such clearly excessive reserve accumulation is China since around 2000. China does have extensive capital controls that block this equilibrating flow of capital. The primary reason that the Chinese government has not taken sufficient adjustment actions to slow or reverse China's balance of payments surplus and accompanying reserves accumulation is that in the short run the needed adjustments would hurt some sectors of the economy, and the government is greatly concerned about the short-run political and social instability that such disruptions might generate. Thus, although adjust-

ment is in China's longer-run economic interest, in the government's calculation, this is outweighed by the likely short-run political costs. Numerous policy announcements have made clear that the government would be happier if the surplus were reduced, but its limited actions reveal the high priority given to avoiding short-run adjustment costs.

The limited adjustments taken by the Chinese government have also contributed to global economic imbalances and the risk they pose to global financial stability. Of course, China alone is not responsible for such global imbalances. From a global perspective, however, the creation of the SWF reduces the national economic cost of these continual reserve accumulations and hence reduces somewhat the pressures on China to do its part toward the needed mutual adjustments.

The greatest concerns about SWFs focus on a different type of issue, however: the fear of foreign government influence over the operation of key segments of domestic economies. As long as SWFs limit themselves to passive portfolio investments they do not pose major problems on this score. The problems come when they make direct investments of substantial stakes in sensitive industries. Such fears have been highlighted by the recent efforts of government-controlled enterprises in China and Dubai to purchase a U.S. oil company and port operator, respectively. Although in neither of these cases was an SWF involved (the bidders were state-owned companies), they have led to considerable speculation about what could occur. Concerns have also been raised that the large amounts of money controlled by some SWFs could generate disruptions in financial markets. These concerns and questions have been fuelled further by the purchase of fairly large stakes in major U.S. financial institutions by various SWFs. Although creating such disruptions would seldom, if ever, be in the interest of SWFs, this may not be sufficient to calm all fears. The likely greater danger is that this range of fear would stimulate protectionist backlashes that could hurt both the SWFs and potential investment recipient countries.

Certainly, some SWFs follow policies that are unlikely to generate conflicts. Norway, for example,

provides considerable transparency and spreads its investments over a wide range of equities, taking only small stakes in any one. Many funds are currently quite opaque, however. Such considerations suggest that the development of an international code of conduct for SWFs could be in the mutual interests of both capital-exporting and capital-importing countries. The possibility of developing such codes has become a major topic of attention in many countries and international forums such as the G7 and the IMF (for instance, see Truman 2007).

See also capital controls; capital flows to developing countries; exchange rate volatility; fear of floating; foreign exchange intervention; global imbalances; hedge funds; International Monetary Fund (IMF); international reserves; reserve currency

FURTHER READING

Lyons, Gerard. 2007. "State Capitalism: The Rise of Sovereign Wealth Funds." Standard Chartered Bank (October 15). United Kingdom. Offers a detailed analysis of SWFs worldwide.

Truman, Edwin, M. 2007. "Sovereign Wealth Funds: The Need for Greater Transparency and Accountability." Policy Brief No. PB07 6 (August). Washington, DC: Peterson Institute. Offers an overview of SWFs, focusing on their lack of transparency, and presents a framework to help evaluate the extent of transparency and accountability of the SWFs.

**THOMAS D. WILLETT
AND RAMKISHEN S. RAJAN**

◼ special and differential treatment

Special and differential treatment (SDT) refers to provisions in World Trade Organization (WTO) agreements under which developing countries or groups thereof receive more favorable treatment or undertake less onerous commitments than other members. Developing countries have long argued that their development status requires that they be subject to different and more favorable trade rules than other members. This principle was accepted in the General Agreement on Tariffs and Trade (GATT, Part IV) and later in the WTO agreements, which contain numerous provisions for SDT for developing countries, as well as additional provisions containing even more favorable treatment for the least-developed countries (LDCs), a group of 50 countries that meet certain development criteria defined by the United Nations.

The main conceptual premise that underlies SDT is that developing countries are intrinsically disadvantaged in their participation in international trade and multilateral agreements involving them, and developed countries must account for this weakness in specifying their rights and responsibilities. A related premise is that trade policies that maximize sustainable development in developing countries differ from those that do so in developed economies, and hence policy disciplines applying to developed economies should not apply to developing countries. The final premise is that it is in the interest of developed countries to assist developing countries in their fuller integration and participation in the international trading system.

The provisions introduced in the WTO agreements fall into two broad categories: positive actions by developed country members or international institutions and exceptions to the overall rules contained in the agreements. Developed countries have also agreed to take three kinds of actions to support developing countries' participation in international trade:

- Provide preferential access to their markets, such as through the Generalized System of Preferences (GSP), under which many developing country manufacturing exports enjoy duty-free entry into developed country markets.
- Provide technical and other assistance to permit them to meet their WTO obligations and otherwise enhance the benefits developing countries derive from international trade.
- Implement the overall agreements in ways that are beneficial or least damaging to the interests of developing countries and LDCs.

Developing countries and LDCs have accepted differential obligations under the WTO agreements, in that they are permitted to undertake policies that limit access to their markets or support domestic producers or exporters in ways not allowed to other members. Examples include GATT Article XVIII on government assistance and the general exemption from reciprocity in trade negotiations with developed countries to reduce or remove tariffs and other barriers to trade. Similar provisions for nonreciprocity are included in the General Agreement on Trade in Services (GATS, Article XIX:2), which states, "There shall be appropriate flexibility for individual developing countries members for opening fewer sectors, liberalizing fewer types of transactions, progressively extending market access in line with their development situation." Second, developing countries and LDCs get more time to meet obligations or commitments under the agreements (for example in Trade-Related Intellectual Property Rights TRIPS).

Both the premises and the practice of SDT have proved controversial. Some have argued that the freedom provided to developing countries to protect their economies has harmed rather than promoted development and, similarly, that the lack of reciprocity in the mutual reduction of tariff barriers has contributed to the relatively high tariff barriers now facing developing countries in developed country markets. It is also argued that the GSP has provided significant benefits to only a few countries and has done little for the poorest and the LDCs.

Others have noted that most of the WTO commitments for SDT made by the developed countries have been vague and legally unenforceable. Many poorer developing countries have been pushed to participate in agreements such as TRIPS, sanitary and phytosanitary measures, technical barriers to trade, and customs valuation without taking into account their lack of capacity to implement their commitments and without ensuring the provision of adequate assistance. Moreover the transition periods envisaged under the WTO were unrealistic. The time limits for extensions had passed, and as of early 2008 there was little evidence that countries had made

sufficient progress in institution building to permit them to implement their obligations. Finally, there is the question of which countries should be receiving SDT. Many low-income and vulnerable economies face the same kind of developmental constraints as the LDCs, yet only the latter receive special consideration, probably because their total participation in world trade is so small as not to constitute a competitive threat to developed country producers. On the other hand, because the principle of self-selection is used in determining who is a developing country, large countries that are well integrated in the international trade such Brazil, China, or Singapore are in principle eligible for the same kind of SDT as small vulnerable economies such as Ghana or St. Lucia.

The Doha WTO Ministerial Declaration underlying the Doha Round of multilateral trade negotiations was replete with pronouncements about SDT. It stated that "provisions for special and differential treatment are an integral part of the WTO agreements," and it called for a review of SDT provisions with the objective of "strengthening them and making them more precise, effective and operational" (para. 44). As of early 2008, little progress had been made on these issues with the exception of a decision related to TRIPS and pharmaceuticals and a number of decisions related to LDCs. It was agreed that LDCs would not have to make any tariff reduction commitments as part of the multilateral negotiations. This was controversial as it raises the risk that LDCs will continue to maintain significant protective barriers that would undermine their development.

In TRIPS, developing countries can use compulsory licensing to reduce the cost of drugs for HIV/AIDS whose patents were held by developed country pharmaceutical companies. But this provision is of no use to countries that have no capacity to produce these drugs domestically. Under heavy public opinion pressure, the WTO agreed to an amendment permitting developing countries that do not have capacity to produce drugs needed to combat epidemics to import them from low-cost suppliers at low prices under carefully circumscribed circumstances. In a related action, negotiators also agreed to

extend the transition period for the implementation of the TRIPS agreement for LDCs until 2013 and for pharmaceuticals until 2016. But the decision does not cover other low-income countries that face similar developmental constraints.

In parallel with the Doha Round negotiations, an effort was made to strengthen the international community's efforts to provide trade-related technical and other assistance. Agreement was reached for an enhanced Integrated Framework for Trade-Related Assistance to LDCs with a new independent secretariat housed in the WTO and the expectation of substantially increased aid resources. On the other hand, a WTO Task Force on Aid for Trade, which was supposed to address assistance needs of all developing countries, could only come up with general and vague recommendations.

Recent experience with SDT suggests an emerging consensus that SDT should be extended to many developing countries that do not have the institutional capacity to implement a number of WTO agreements — or for whom implementation of such agreements is not a development priority — and that many of these countries, especially the LDCs, are deserving of increased trade-related international assistance. Also, a number of proposals to strengthen monitoring of SDT needs and implementation, possibly in the context of WTO trade policy reviews, have received widespread support, though no specific agreement has been reached.

Controversies are likely to continue on specific developing country SDT proposals, however, especially those that affect basic WTO disciplines such as binding of all tariffs, commitments not to use nontariff barriers to trade, or participation in the multilateral trade negotiations, as well as on which countries should receive SDT. Overall, the role of SDT in the world economy is going to continue to be small but important to a considerable number of poor developing countries, especially the LDCs.

See also access to medicines; Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS); Doha Round; Trade Policy Review Mechanism; trade-related capacity building

FURTHER READING

Hoekman, Bernard, Constantine Michalopoulos, and Alan Winters. 2004. "Special and Differential Treatment of Developing Countries in the WTO." *World Economy* 27 (4): 481 506. A recent assessment of the economic merits of various approaches to SDT.

Hudec, Robert. 1987. *Developing Countries in the GATT Legal System.* Hampshire, UK: Gower. Basic reference to the legal provisions pertaining to SDT in the GATT, which were later incorporated in the WTO.

Michalopoulos, Constantine. 2001. *Developing Countries in the WTO.* Houndmills, Hampshire, UK: Palgrave. Contains a discussion of the merits of SDT in the context of trade theory and experience with developing country participation in the WTO.

World Trade Organization. 2000. "Implementation of Special and Differential Provisions in WTO Agreements and Decisions." WT/COMTD/W/77. Geneva: WTO. Basic reference that contains a listing of all the SDT provisions in WTO agreements prepared in advance of the Doha Round.

CONSTANTINE MICHALOPOULOS

■ special drawing rights

Special drawing rights (SDRs) are an internationally recognized unit of account and reserve assets issued by the International Monetary Fund (IMF) and allocated to IMF member countries in proportion to their quotas at the IMF. According to the IMF (2006a), the SDR is "a potential claim on the freely usable currencies of IMF members." SDRs can be exchanged for other currencies in two ways: first, through voluntary exchanges between members; and second, by an IMF directive, designating members with strong external positions to purchase these SDRs.

SDRs were initially created in 1969 to increase the availability of easily convertible reserve assets. Before 1969, reserve assets — those assets held by central banks to clear international transactions — were held mostly in U.S. dollars and gold. The economists Peter Clark and Jacques Polak (2004) note that, at the time, issuing SDRs was seen as a way of preventing

official dollar holdings from undermining the stability of the system. Reserve accumulation was perceived as destabilizing since many central banks converted their dollar reserves into gold, thereby drawing down the limited U.S. gold stocks.

SDRs are also used as units of account by the IMF in all of its transactions. The IMF conditional loans are denominated in SDRs, while the interest rate on these loans is calculated on the basis of the SDR interest rate (in addition to an interest surcharge that varies with the different lending facilities).

Amended every five years, the value of an SDR unit is the sum of the values of the following as of January 2006: 0.632 U.S. dollars, 0.410 euros, 18.4 Japanese yen, and 0.0903 British pounds. As of February 1, 2007, 1.496137 U.S. dollars were worth one SDR unit. The SDR interest rate is calculated as the weighted average of interest rates on short-term instruments in the financial markets of the four currencies included in the SDR valuation basket; it is posted on the IMF Web site once a week. On the week of February 1, 2007, for example, the SDR interest rate was 4.20 the weighted average of the interest rates of a three-month U.S. Treasury bill, a three-month Europe rate, the Japanese government's 13-week financing bill, and a three-month UK Treasury bill.

Role of SDRs From their onset, SDRs could be held only by governments, central banks, and official bodies such as the IMF. The original SDR allocation of 9.3 billion SDRs was disbursed among the IMF members according to predetermined quotas over a two-year period (1970 72). An additional amount of 12.1 billion SDRs was allocated and distributed between 1979 and 1981. SDRs have not been allocated since. As a result of the reluctance among the IMF stakeholders to allocate new SDR issues, SDRs now amount to only about 1 percent of the international reserves held worldwide. SDRs clearly did not end up as the primary reserve asset in the global monetary system as envisioned in the IMF Article of Agreements (XXII). In addition, Clark and Polak (2004) suggest that the concerns of international liquidity (i.e., the worldwide lack of sufficient gold reserves) that led to the establishment of the SDR no longer apply.

Although the original rationale for the creation of SDRs is no longer relevant to the post Bretton Woods system of flexible exchange rates, other reasons for their existence as reserve assets remain. Since capital accounts are now much more open than ever before (with the possible exception of 1880 1914), reserves are seen as a means to prevent fluctuations in the trade balance and domestic consumption that would be required in the face of fluctuations in capital flows. SDRs enable countries to diversify their reserve holdings, and more important for developing countries, the SDR can be held at a much lower cost than major currencies such as the U.S. dollar, the Japanese yen, the euro, or the British pound. This is especially true for the lowest-income countries, which are virtually cut off from the international financial markets and for whom the only means of obtaining reserves is running trade surpluses through reductions in imports.

Another possible future use for the SDR is as an alternative reserve asset that can be issued broadly in case of a dramatic U.S. dollar crash (Lissakers 2006). A dollar crash, a plausible event in light of the persistent U.S. balance of payment deficits, and the central role of the dollar as a reserve currency (roughly 70 percent of all foreign reserves are held in U.S. dollar denominated assets), might create the conditions for a general flight from the dollar; this means that the question of constraints on international liquidity might reemerge.

See also balance of payments; Bretton Woods system; dollar standard; dominant currency; global imbalances; gold standard, international; hot money and sudden stops; international liquidity; International Monetary Fund (IMF); international reserves; reserve currency; Triffin dilemma; twin deficits; vehicle currency

FURTHER READING

Clark, Peter B., and Jacques J. Polak. 2004. "International Liquidity and the Role of the SDR in the International Monetary System." *IMF Staff Papers* 51 (1): 49 71. A

discussion of the 1969 70 introduction of SDR, written by senior IMF staffers at the time who argue for a re invigoration of the SDR allocation program.

Goldstein, Henry N. 1969. "Gresham's Law and the De mand for NRUs and SDRs." *The Quarterly Journal of Economics* 83 (1): 163 66. A contemporary argument supporting the introduction of SDRs.

International Monetary Fund. 2006a. "A Factsheet Special Drawing Rights (SDRs)." Downloadable from http://www.imf.org/external/np/exr/facts/sdr.htm (ac cessed August 15, 2006). Contains updated information about the current status and history of SDRs.

International Monetary Fund. 2006b. "SDR Valuation." Downloadable from http://www.imf.org/external/np/ fin/rates/rms sdrv.cfm (accessed August 17, 2006). Contains daily updates on the value of SDRs.

International Monetary Fund. 2006c. SDR Interest Rate Calculation. Downloadable from http://www.imf.org/ external/np/fin/rates/sdr ir.cfm (accessed August 17, 2006). Contains details on the current SDR interest rate.

Lissakers, Karin. 2006. "Is the SDR a Monetary Dodo? This Bird May Still Fly." In *Reforming the IMF for the 21st Century*, edited by Edwin M. Truman. Washington, DC: Institute for International Economics, Special Re port #19. Written by a former U.S. executive director at the IMF, the paper advocates the allocation of SDRs in the case of a U.S. dollar crash. The book includes other useful discussions on various aspects of IMF reform proposals.

ILAN NOY

■ special economic zones

See South-South trade

■ specific-factors model

Most undergraduates in their first year in economics have been exposed to an important ingredient in the specific-factors model, a model in which each in dustry employs some factor used only in that par ticular industry. What happens to total output when more of a variable factor is added in a production process to a fixed quantity of this specific factor? The answer exemplifies the Law of Diminishing Returns, and this ingredient is basic in the general equilibrium context in which the specific-factors model is set. David Ricardo used this concept when referring to the differential rents that various qual ities of land would receive, and Viner (1931) made use of it in his famous article in which he argued with Wong, his draftsman. Haberler (1936) in his classic text on international trade presents verbally some of the logic of what Samuelson (1971) later referred to as the Ricardo-Viner model. The formal exposition of the model in general equilibrium terms was carried out by Samuelson and by Jones (1971), both of whom emphasized the usefulness of the model in the theory of international trade, although the model itself is also applicable to closed economies.

The nature of the model is most easily discussed in the context in which only a pair of commodities is produced in competitive conditions, with each commodity making use of a factor employed only in that sector (i.e., a *specific* factor) as well as a factor of production (typically taken to be labor) that is used in both sectors (the *mobile* factor). In a competitive equilibrium, factor prices and input-output coeffi cients adjust to maintain full employment of all factors, and with constant returns to scale charac terizing production processes, costs of production adjust to equal commodity prices if both commod ities are produced. The process of solving the model formally for changes in factor prices and commodity outputs when commodity prices or factor endow ments are altered is more simple than that found in the Heckscher-Ohlin model (the standard model used in trade theory since the Stolper-Samuelson article appeared in 1941) in that it is not necessary to solve more than one equation at a time. The key equation is the one that asserts full employment of the mobile factor. With techniques in each sector depending only on the wage rate relative to the commodity price in that sector (the important technological parameter being the elasticity of the

demand curve for labor, exhibiting diminishing returns to labor as more is added to a given amount of the specific factor) and each output restricted by the given amount of the specific factor and the intensity of its use (which depends on the wage/price ratio), the change in the wage rate is seen to depend on commodity price changes and changes in factor endowments. These relationships are often portrayed in a "back-to-back" diagram, such as figure 1. A pair of (value of) marginal product schedules face each other, each assuming given values for the quantity of the specific factor available in that sector as well as that sector's commodity price. The intersection point, A, reveals both the equilibrium value of the wage rate and the quantity of labor assigned to each sector.

Details of the solution are found in many places, for example, in the supplement to chapter 5 of the Caves, Frankel, and Jones text (2007). The important results are (1) if either commodity price increases, the wage rate also increases, but by less than in proportion; (2) any increase in the labor endowment at constant commodity prices drives down the wage rate (to the benefit of both specific factors); and (3) any increase in the endowment of either specific factor lowers that factor's return, pushing up the nominal wage rate, and thus driving down the return to the other specific factor as well. The asymmetry found in factor returns when the price of a single commodity increases reflects the asymmetry found in the mobility of the two factors: an increase in the price of the first commodity must lead to a matching increase in average cost. The return to labor is constrained by its use as well in the other sector, one that has not benefited by a price increase, while the return to the specific factor used only in the first sector is not constrained in this fashion. The consequence is that the wage rate cannot increase relatively as much as p_1, thus pushing up the return to the specific factor employed there by a magnified relative amount so that unit costs increase as much as price. With the nominal wage rate rising, the return to the specific factor in the second sector falls.

An important result in the field of political economy is immediately apparent, even in the two-commodity setting. Suppose the primary issue facing voters before an election is whether or not to impose a tariff on imports. The specific factor in the import-competing sector will be strongly in support of the tariff, while the other specific factor would be strongly opposed. This is not surprising. But what of the attitude of voters whose income is in the form of wages? Protection raises the nominal wage rate but also increases the cost of living. This may leave many voters with the desire to stay at home on election day (especially if the weather is inclement) since their real incomes do not depend that heavily on the outcome. (Ruffin and Jones 1977 argue that on balance labor will be mildly against protection in the specific-factors model.) This would help to explain relatively low voter turnout at election times in countries such as the United States. If, instead, the important issue before the voters concerns immigration, both specific factors stand to gain or lose together, and this may result in their joining forces in a political alliance.

The setting of the specific-factors model has two basic interpretations. On the one hand, the two specific factors may fundamentally be different say, land and capital (e.g., in Jones 1971). On the other hand, they may represent, say, two kinds of capital that are specific in the short run but can become interchangeable with the passage of time (e.g., in Neary 1978). As Magee (1980) has argued, a
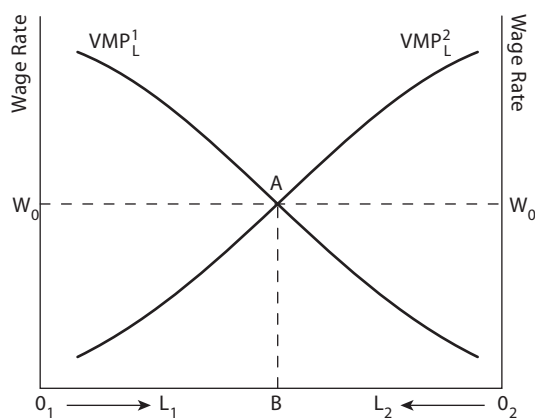
Figure 1
Wage rate determination

sector-specific type of capital might change its attitude toward protection in the long run. The Neary interpretation has become popular in explanations of how the specific-factors model may be linked to the standard Heckscher-Ohlin model, where both factors are intersectorally mobile. Such a link was given a different rationale in the model of Sanyal and Jones (1982), in which a country produces final commodities by using labor and *middle products,* that is, goods in process, raw materials, or intermediate goods that can be obtained on world markets. The country may export some of its own production of middle products in exchange for imports that are better suited to its own needs in producing final consumption goods. Thus middle products produced at home with labor and the country's own specific factors can be traded for middle products requiring, say, specific factors not available at home. In other words, final consumer goods are produced with two mobile factors: labor and traded middle products. (As the Canadian economist Doug Purvis once remarked, in this setting Heckscher-Ohlin does not explain trade, trade explains Heckscher-Ohlin.)

How does the specific-factors model match up with the Heckscher-Ohlin model? In the setting in which only two commodities are produced, differences between the specific-factors model and the Heckscher-Ohlin model are often emphasized in the theory of international trade, especially as regards the effects of free trade on a nation's factor returns. As Samuelson (1948) demonstrated, if endowment differences are relatively small between two countries sharing the same technology and facing the same traded goods prices, factor prices tend to be equalized by trade despite the fact that each factor has a purely national market. The specific-factors model does not share this property. With the number of factors (3) exceeding the number of produced commodities (2), any tendency for factor returns to become equalized with trade disappears. A related comparison concerns the effects of factors becoming mobile between countries. In the two-factor, two-commodity Heckscher-Ohlin model with countries sharing the same technology and endowments not too dissimi-

lar, a movement of factor(s) from one country to another can be absorbed by a change in the composition of outputs without requiring any change in factor prices. Not so in the specific-factors model with three factors and two commodities, because at given commodity prices changes in factor endowments exercise a direct effect on factor returns. This latter model is often more appreciated by labor economists, who may expect labor immigration to have a depressing effect on national wage rates.

The difference between these two models tends to be less apparent when a specific-factor model with n commodities (and $n+1$ factors, only one of which is mobile) is compared with the so-called *strong* form of the n-factor, n-commodity Heckscher-Ohlin model in which the increase of any commodity price serves to raise the return to the factor used relatively intensively in that sector and to lower the returns to all other factors (e.g., see Kemp and Wegge 1969). Such a *magnification effect* of commodity prices on factor returns is shared by the specific-factors model   the return to one specific factor is raised, and to all others is lowered. Only the return to mobile labor is not so magnified. Indeed, suppose the mobile factor is not labor, but rather some intermediate good that is produced by all of the "specific" factors. This *produced mobile factor* model (Jones and Marjit 1991) becomes a Heckscher-Ohlin model that satisfies the strong conditions   one real winner and all others losers when a price changes   cited earlier.

The case can be made that the specific-factors model   with each sector having a unique specific factor but sharing mobile labor with all other sectors   is especially useful as a general equilibrium model of production because it does generalize so readily to higher dimensions and has highly appealing qualities: (1) If a single commodity price increases, the output of that commodity also rises, drawing resources (mobile labor) from *all* other sectors; (2) if the endowment of a specific factor increases (and commodity prices remain constant) not only is its return lowered, but the consequent increase in the wage rate pushes down the returns to *all* other specific factors. This latter result may yield a surprising consequence   the returns to some other

specific factors may fall by a greater relative extent than does that of the factor whose endowment has increased.

Although the Heckscher-Ohlin model may not be especially useful in the many-factor case because extremely detailed structure must be imposed before explicit solutions can be obtained, it does have a distinct advantage in that the two-factor scenario is consistent with a world in which trade takes place in many commodities. There is no two-factor, many-commodity version of the specific-factor model. The theory of international trade emphasizes that trade encourages each country to specialize in a few activities in which it has the greatest comparative advantage, and the two-factor Heckscher-Ohlin model can well illustrate that *which* commodities a country produces depends both on world commodity prices and local factor endowments. This production choice becomes endogenous   it can vary as, say, the country grows and the capital/labor endowment ratio expands. Suppose, however, that a commodity that had been produced in the past now cannot earn as high a return on its capital as some new commodity on the trading scene. If this capital had become specific, that industry may nonetheless stay in business if its capital can earn anything exceeding scrap value. That is, the specific-factor model may prove useful in modeling why it is that some industries still produce even if they would not be viable if new capital had to be raised for production (Jones 2007).

Some disenchantment with general equilibrium models seems to be based on the widespread difficulty of obtaining comparative static results when an original equilibrium is disturbed by some change in prices or endowments. Detailed structure must be imposed on the model. Sufficient structure is a characteristic of specific-factor models, and further higher-dimensional results can be obtained by employing the following kind of *variation* on such models. Consider first the way in which a "bubble" diagram can be used to illustrate a three-factor, two-commodity specific factors model in figure 2(a). Each of the two X_i outputs is produced with specific factor K_i and mobile labor, L. Figure 2(b) adds another two sectors, with notations suggesting two

countries, with Y-type capitals, K_Y^* and K_Y, specific both to country and occupation. L^* and L are, respectively, foreign and home labor forces, each specific to the country but mobile between sectors, and K_X, assumed specifically used in the X industry, is now internationally mobile. This is a five-factor, four-commodity model, whose structure has specific factors used only for the "end" products and three types of mobile factors   labor in each country mobile between sectors but not internationally and X-type capital, mobile between countries but sector-specific. It is not difficult to analyze (see Jones 2000, chapter 3) because it can be treated in two stages. In the first, suppose that the allocation of X-type capital between home and foreign industries is kept the same so that figure 2(b) resembles two countries, each of the figure 2(a) type. In world markets suppose the prices of X-type goods increase by the same relative amount and that of Y-type goods stay the same. In each country separately standard specific-factor results are obtained: the return to X-type capital increases relatively more than does the price of X goods, the wage rate rises but by less than X's commodity price, and the return to Y-type capital falls. In the second stage, let X-type capital become mobile internationally. In which direction does it flow? Toward the home country if, and only if, in stage 1, the return to X-type capital at home increases by more than it does abroad.

Such a comparison depends largely, and indirectly, on how much the wage rate is stimulated in each country, and this comparison, in turn, is revealed by the solution for wage changes in the
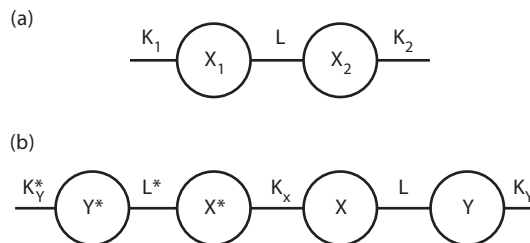
Figure 2

Bubble diagram of three factor, two commodity specific factors model and five factor, four commodity specific factors model

specific-factors model. The formal solution was not derived earlier, but it can be shown (e.g., in Caves, Frankel, and Jones 2007; Jones 2000, chapter 3) to depend on the product of three terms: s_X, i_X, and θ_X. s_X is the relationship between the elasticity of demand for labor (i.e., of the marginal product curve) in the X sector compared to the economy average. If X is "typical," this has value unity; i_X is equal to labor's distributive share in the X sector relative to its share in the economy and would be greater than unity if and only if X is relatively labor intensive. Finally, θ_X is the fraction of the country's income devoted to the production of X. If the X sector is fairly typical in terms of labor demand elasticity and labor intensity, everything depends on the relative size of the X industry. Suppose this is larger at home than abroad. If so, the wage rate will tend to increase more at home than abroad, and since both countries experience the same price rise for X, the return to capital will increase in both countries but tend to increase relatively more abroad. With a greater increase in the wage rate at home, fewer rents are available to attract capital to the home country.

This two-stage process illustrates how the variation in the specific-factors model that is exemplified for the four-commodity case in figure 2(b) leads to a modeling strategy that makes use of the specific-factor logic at the first stage and then completes the analysis by asking how rates of return to a mobile factor compare in two countries. Jones and Marjit (2003) explore a different interpretation of figure 2(b) in which the two labor forces shown there correspond to a single country's supply of skilled and unskilled labor. An interesting question might concern the consequences for the wage premium and wage levels of a training program that converts some unskilled labor into skilled labor.

The specific-factors model is a simple form of general equilibrium model in which an extreme asymmetry in factor mobility is assumed—one factor is mobile and the others are not. A model in which all factors have partial mobility would of course be more complex, even if attractive for modeling dynamic movements. The Heckscher-Ohlin model also has extreme assumptions—namely, that the two factors

(in the two-factor case) have the same degree of mobility instantaneously. As all theorists acknowledge, simplicity in model building is a strong virtue unless it rules out features of the setting that are of most concern. As illustrated earlier, departures from the specific-factors setting can often be better understood by making use of embedded features that do correspond to the specific-factor model even if other features are also involved. And, as Samuelson and others have often remarked, it is the general equilibrium model that captures much of the basic reasoning about diminishing returns in partial equilibrium settings.

See also comparative advantage; Heckscher-Ohlin model; migration, international; political economy of trade policy; trade and wages

FURTHER READING

Caves, Richard, Jeffrey Frankel, and Ronald W. Jones. 2007. *World Trade and Payments*. 10th ed. Cambridge, MA: Addison Wesley. A standard textbook used in undergraduate courses.

Haberler, Gottfried. 1936. *The Theory of International Trade*. London: Wm. Hodge. A classic prewar exposition of the theory of trade.

Jones, Ronald W. 1971. "A Three Factor Model in Theory, Trade, and History." In *Trade, Balance of Payments, and Growth*, edited by Jagdish Bhagwati, Ronald Jones, Robert Mundell, and Jaroslav Vanek. Amsterdam: North Holland, chapter 1. Along with Samuelson (1971) the pair of articles setting out the algebraic exposition of the specific factor model.

———. 2000. *Globalization and the Theory of Input Trade*. Cambridge, MA: MIT Press. Discusses how international trade theory is altered when some inputs into production have international markets.

———. 2007. "Specific Factors and Heckscher Ohlin: An Intertemporal Blend." *Singapore Economic Review* 52: 1 6. Points out how a blend of the two factor, many commodity version of Heckscher Ohlin theory, combined with the $(n+1)$ by n dimensional specific factors model, provides a simple model of competitive trade allowing both a heavy degree of specialization with trade and a variety of products produced by a given country.

Jones, Ronald W., and Sugata Marjit. 1991. "The Stolper Samuelson Theorem, the Leamer Triangle, and the Produced Mobile Factor Structure." In *Trade, Policy, and International Adjustments*, edited by A. Takayama, M. Ohyama, and H. Ohta. San Diego: CA: Academic Press. An exposition of a simple *n* factor, *n* commodity model.

———. 2003. "Economic Development, Trade, and Wages." *German Economic Review* 4: 1 17. An adapta tion of figure 2 to the question of the wage premium between skilled and unskilled workers.

Kemp, Murray, and Leon Wegge. 1969. "On the Relation between Commodity Prices and Factor Rewards." *In ternational Economic Review* 9: 497 513: An extension of the Stolper Samuelson theorem to three factors and three commodities.

Magee, Stephen. 1980. "Three Simple Tests of the Stolper Samuelson Theorem." In *Issues in International Eco nomics*, edited by P. Oppenheimer. London: Oriel Press, 138 53: An argument for the superiority of the specific factors model in explaining supports for commercial policy.

Neary, J. Peter. 1978. "Short Run Capital Specificity and the Pure Theory of International Trade." *Economic Journal* 88: 488 510. Shows how the passage of time can convert a specific factors setting to the all factors mobile Heckscher Ohlin setting.

Ricardo, David. 1817. *The Principles of Political Economy and Taxation*. Reprint, 1981. Cambridge: Cambridge University Press. Shows both the argument for com parative advantage and the analysis of rents in a specific factor setting.

Ruffin, Roy, and Ronald W. Jones. 1977. "Protection and Real Wages: The Neo Classical Ambiguity." *Journal of Economic Theory* 14: 337 48. Establishes a presumption that workers would prefer free trade to protection.

Samuelson, Paul A. 1948. "International Trade and the Equalisation of Factor Prices." *Economic Journal* 58: 163 84. The classical argument that factor prices could be equalized between trading countries with different factor endowments.

———. 1971. "Ohlin Was Right." *The Swedish Journal of Economics* 73: 365 84. An exposition of the algebra of the specific factor setting.

Sanyal, Kalyan, and Ronald W. Jones. 1982. "The Theory of Trade in Middle Products." *American Economic Re view* 72: 16 31. An analysis of trade that is in the "middle" of the production spectrum.

Stolper, Wolfgang, and Paul A. Samuelson. 1941. "Pro tection and Real Wages." *Review of Economic Studies* 9: 58 73. Shows how, in a Heckscher Ohlin setting, a country would raise real wages if it imposes a tariff on labor intensive imports.

Viner, Jacob. 1931. "Cost Curves and Supply Curves." *Zeitschrift für Nationalökonomie* 3: 23 46. Reprint, 1953. American Economic Association, *Readings in Price Theory*. London: Allen and Unwin. A discussion of a firm's short run and long run average cost curves, with a dispute over geometry.

RONALD W. JONES

■ speculation

The economist John Maynard Keynes (1936) de fined speculation as the purchase of securities at a price above their fundamental value with a view to sell them at yet a higher price in the subsequent trading periods. Ever since the economist Milton Friedman (1953) argued that such destabilizing speculation would be unprofitable and thus unsus tainable in the long run, economists generally have discounted the possibility that speculation could cause asset prices to deviate from fundamental values, that is, prices warranted by the *true* earning potential of firms. The intuition behind Friedman's argument rested on a simple view of arbitrage. In a market that includes both smart traders who know the true values and those who make whimsical and misinformed trades, generally referred to as noise traders, the latter create riskless arbitrage opportunities from which smart traders could profit.

The rise of behavioral finance theory has made economists more receptive to studying the limita tions of arbitrage in real-world markets. Informed traders with a "short" trading horizon who sell overvalued assets short can find that by the time they are supposed to close their position, the true value has increased, or the assets in question have become even more overpriced. Because the smart traders who have sold securities short could make losses in either sit-

uation, they limit the initial positions they take in an over- or undervalued asset, preventing the current price from smoothly adjusting to its true value as Friedman had envisioned. It might even be the case that it pays for smart traders to act like noise traders themselves in the short run, bidding prices further away from true values rather than helping to close the gap between the two (De Long et al. 1990; Griffin, Harris, and Topaloglu 2003). Thus an increasing number of economists now hold that *riskless* arbitrage is not always effective (Shleifer 2000).

This view is reminiscent of Keynes's (1936, chapter 12) famous beauty contest analogy, where speculators base their expectations of future asset prices not only on what they think the true value is, but, more important, on what they think the average opinion about the average opinion is. In other words, *noise* is at least as important as information about true values in causing asset price changes, rendering the resale price uncertain (Black 1986). This, in turn, implies that traders must not only form higher-order expectations (i.e., on what others think others think) but also decide how much weight to assign them relative to what they themselves think the true value is (Hirota and Sunder 2003).

If a trader observes that the price of an asset (or an asset group) which she thinks is already overvalued is still rising in price, she is led to surmise that either her opinion about the true value is wrong or that the price is rising on account of market sentiment (Abreu and Brunnermeier 2003). In either case, the information and opinion of others, as revealed in current price changes, are likely to gain in importance in how the trader forms his/her expectation about the future price. Such information becomes either a proxy for the higher-order expectations or a corrective on opinions about the true value, or some combination of both. Under these conditions whether speculation is stabilizing or not crucially depends on the relative weight traders assign to their higher-order expectations (i.e., what they think others think others think) relative to their own assessment of what the true value is. Though it appears to be long forgotten, this basic idea in its simplest form goes back to Kaldor (1939), where whether speculation is stabilizing or not de-

pends on the elasticity of future price expectations with respect to current price changes. In this early formulation, stability requires a less than unitary elasticity of expectations, where traders revise their expected future price proportionally less than the change in the current price.

In the modern literature, much more elaborate analyses of cases where agents make trades relying on information revealed by the past decisions of other traders rather than their own abound and are generally discussed under the heading of *herding* behavior. These generally model "momentum-investor" or "positive-feedback" strategies, where traders tend to buy assets whose price has been on the rise and sell those that have been falling in price. Short trading horizons, the sequential nature of trades, and information costs are the real-world market attributes that are emphasized in this body of work.

In contrast to standard models of asset pricing that implicitly assume long-term horizons, short-term trading horizons play an important role in persistent deviations of asset prices from true values (Dow and Gorton 1994). In these models, it is argued that the incentive structure that defines money managers' employment gives rise to agency problems (Allen and Gorton 1993), making it rational for speculators to have short trading horizons. In a world characterized by market imperfections and uncertainty, speculators who tie their resources to long-term investments would fail to exploit profitable investment opportunities that would unexpectedly arise (Shleifer and Vishny 1997).

Second, in real-world markets trading is sequential, which implies that traders observe and can learn from one another. Early action can enable traders to rationally exploit information revealed by the actions of other agents. Profit-maximizing traders can thus successfully focus on what other traders also know rather than trying to learn information others do not have. A narrow set of information can then become the primary focus of attention even when it has little bearing on fundamentals (Scharfstein and Stein 1990).

Finally, especially in international markets, gathering information requires large fixed costs which

generate economies of scale for large investors. Thus the greater the cost of acquiring information, the higher is the incentive for uninformed small traders to imitate large investors (Calvo and Mendoza 2000). This also causes small investors to be much more aggressive, especially in their selling in markets where large investors have a significant presence (Corsetti et al. 2001), giving rise to increased volatility and likelihood of abrupt shifts in capital flows, leading to currency and financial crises.

See also contagion; currency crisis; exchange rate volatility; financial crisis; spillovers

FURTHER READING

Abreu, D., and M. Brunnermeier. 2003. "Bubbles and Crashes." *Econometrica* 71: 173 204. Shows that a bubble can exist despite the presence of rational arbitrageurs.

Allen, F., and G. Gorton. 1993. "Churning Bubbles." *Review of Economic Studies* 60: 813 36. Argues that agency problems caused by asymmetric information in financial markets cause financial assets to trade at prices different than fundamentals.

Black, F. 1986. "Noise." *Journal of Finance* 41 (July): 529 43. Argues that *noise* in financial markets plays a profound role in shaping our views of the world.

Calvo, G., and E. Mendoza. 2000. "Rational Contagion and the Globalization of Securities Markets." *Journal of International Economics* 51: 79 113. Argues that the large fixed costs of gathering information in foreign exchange markets incentivize small traders to imitate large investors.

Corsetti, G., A. Dasgupta, S. Morris, and H. S. Shin. 2001. "Does One Soros Make a Difference? A Theory of Currency Crises with Large and Small Traders." *Review of Economic Studies* 71: 87 113. Argues that the very existence of large investors in foreign exchange markets of emerging countries increases the vulnerability of these countries to speculative attacks.

De Long, J. B., A. Schleifer, L. Summers, and R. Waldmann. 1990. "Positive Feedback Investment Strategies and Destabilizing Rational Speculation." *Journal of Finance* 45 (2): 379 95. Shows that rational speculation might entail rational traders to act as noise traders, jumping on the bandwagon of an asset price bubble the latter might create.

Dow, J., and G. Gorton. 1994. "Arbitrage Chains." *Journal of Finance* 49: 819 49. Explores how and why limited trading horizons can give rise to inefficient asset prices.

Friedman, M. 1953. "The Case for Flexible Exchange Rates." *Essays in Positive Economics*. Chicago: Chicago University Press. In this famous paper, Friedman argued that speculation had to be stabilizing for it to be profitable.

Griffin, J., J. Harris, and S. Topaloglu. 2003. "The Dynamics of Institutional and Individual Trading." *Journal of Finance* 58 (6): 2285 2320. Summarizes the findings of studies that examine the relationship between institutional ownership and stock returns.

Hirota, S., and S. Sunder. 2003. "Price Bubbles sans Dividend Anchors: Evidence from Laboratory Stock Markets." Working Paper. Downloadable from http://www.som.yale.edu/faculty/Sunder/research.html. Tries to model higher order expectations about what the average opinion expects what the average opinion expects and on and on.

Kaldor, N. 1939. "Speculation and Economic Stability." *Review of Economic Studies* 6 (3): 1 27. Specifies the conditions under which speculation can be stabilizing.

Kemp, M. C. 1963. "Speculation, Profitability, and Price Stability." *Review of Economics and Statistics* 45: 185 89. Provides an early criticism of Friedman's argument that profitable speculation must be stabilizing.

Keynes, J. M. 1936. *The General Theory of Employment, Interest, and Money*. Reprint, 1964. New York: Harcourt Brace Jovanovich. Keynes's major theoretical work where he lays out his theory of effective demand.

MacDonald, R. 2000. "Expectations Formation and Risk in Three Financial Markets." *Journal of Economic Surveys* 14 (1): 70 100. An overview of the empirical literature on expectations formation in financial markets.

Scharfstein, D., and J. Stein. 1990. "Herd Behavior and Investment." *American Economic Review* 80 (3): 465 89. Examining the conditions under which herd behavior occurs in financial markets, the paper shows why it can be rational for some traders to mimic other traders rather than relying on substantive private information.

Shleifer, A., and L. Summers. 1990. "The Noise Trader Approach to Finance." *Journal of Economic Perspectives* 4

(2): 19 33. Argues that riskless arbitrage is severely limited not only in relation to the prices of shares or bonds as a whole but also in relation to the prices of individual assets.

Shleifer, A., and R. Vishny. 1997. "The Limits of Arbit rage." *Journal of Finance* 52 (1): 35 55. Discusses how anomalies in financial markets can appear and why ar bitrage would fail to eliminate them.

KORKUT A. ERTURK

■ **spillovers**

In international finance, spillovers are phenomena in which a change in asset price (or asset price volatility) in one country leads to a change in asset price (or asset price volatility) in another country. Interest in this topic resulted from the simultaneous declines in global equity markets during the 1987 stock market crash and the high comovement of asset returns in many countries during the financial crises in the late 1990s. This topic is crucial for our understanding of how information, often proxied by changes in asset price or volatility, in one market transmits to another market. In addition, the effect of foreign asset prices on domestic asset prices has implications for how we should value domestic assets and how we should allocate wealth across different countries.

Empirical Studies Most studies generally examine the spillover effect from three financial centers the United States, the United Kingdom, and Japan to other international financial markets, and, because of data availability, they focus mainly on equity markets. Existing studies use different sample periods, data frequencies, and econometric methods. The standard methods used to measure the spillover effect can be divided into two groups. The first group measures the spillover effect on changes in asset price (asset return). The second group measures the spillover effect on asset return volatility. This method examines the influence of foreign asset return volatility on domestic asset return volatility. Because the direction of the effect of foreign return on domestic return can vary over time, it is easier to measure the volatility spillover effect as changes in asset price

(independent of direction), which are always reflected in increases in asset return volatility; thus, the effect is always positive. The volatility spillover effect can give us only an indirect measure of asset price linkages, however.

These studies' general findings are: (1) return and volatility spillovers are present among major markets; (2) the majority of the spillover effect appears to originate from the United States to other foreign financial markets; (3) when volatility is high, the spillover effect in returns tends to be higher returns are more correlated; and (4) the spillover effect on volatility is asymmetric, with the spillover effect from bad news (negative unexpected return) having larger effects than that from good news (positive unexpected return) (see Gagnon and Karolyi 2006).

Sources of Spillovers Spillovers have several possible sources. First, economic linkages between countries may lead to spillovers. A common shock to global economic fundamentals may lead to an observed spillover effect. For example, a surprise change in U.S. monetary policy stance may affect foreign asset prices simultaneously, which may generate the impression of a spillover effect. Alternatively, a change in an asset price in one country may affect another country's economic fundamentals. For example, a devaluation of one country's exchange rate may affect other countries' international trade, either directly or indirectly, which may in turn affect their asset prices. By and large, the empirical evidence supporting observable macroeconomic linkages as a primary source of spillovers is very weak.

Second, information asymmetry may lead to transmission of country-specific shocks from one country to another. This explanation was popularized following the stock market crash in 1987. This event not only caught market participants by surprise, but it was also hard to attribute the movements to economic fundamental linkages. The economists King and Wadhwani (1990) propose an information-based model in which price changes in one market depend on price changes in other markets and a domestic investor who observes a price change in a foreign market cannot distinguish between changes driven by economic fundamentals

and those driven by foreign country specific shocks (which should not have any influence on domestic asset prices). The implication of the model is that a price change attributable to foreign-specific shocks can lead to price changes in other countries, which in turn will increase volatility in foreign countries and increase the correlation between the two markets. The term for this channel of spillover with no economic fundamental link is *market contagion*.

Third, trading behavior may generate spillover effects. Following the high correlation among global financial markets during the financial crises in the late 1990s and the increase in global cross-border financial investments, global institutional investors have garnered attention as another source of spillover effects. Similar to the second source of spillovers, this source does not rely on economic fundamental linkages among countries. The spillover effect that occurs during financial crises is often referred to as *contagion*. Spillovers can occur because investors need to liquidate some of their assets when they lose money (Kyle and Xiong 2001) or face liquidity constraints (Yuan 2005), and they choose to sell assets in different countries simultaneously. Alternatively, spillovers may be the outcome of cross-market rebalancing (Kodres and Pritsker 2002). *Cross-market rebalancing* refers to investor behavior in which investors respond to changes in asset price in one country by optimally adjusting their asset holdings in other countries.

Although studies agree on the general empirical features of spillovers, their source is still a challenging research topic. Most studies find little evidence to support the role of economic linkages in explaining spillovers, and only a few studies find evidence to support the role of information asymmetry and trading behaviors in explaining spillovers. The lack of consensus on explanations for spillovers reflects the difficulty of measuring their source (e.g., information asymmetry or investors' trading behaviors). Future researchers will want to use new data and new econometric techniques to reexamine current explanations of spillovers before exploring new hypotheses.

See also asymmetric information; banking crisis; capital flows to developing countries; contagion; currency crisis; financial crisis; international liquidity

FURTHER READING

Gagnon, Louis, and G. Andrew Karolyi. 2006. "Price and Volatility Transmission across Borders." *Financial Markets, Institutions, and Instruments* 15 (3): 107 58. A review of the research on the dynamics of asset price and volatility spillovers across international financial markets.

King, Mervyn A., and Sushil Wadhwani. 1990. "Transmission of Volatility between Stock Markets." *Review of Financial Studies* 3 (1): 5 33. A classic, often cited reference to the study of volatility spillovers.

Kodres, Laura E., and Matthew Pritsker. 2002. "A Rational Expectations Model of Financial Contagion." *Journal of Finance* 57 (2): 769 99. A theoretical model that demonstrates the role of investors' optimal trading strategies in generating return comovements among assets not sharing any common economic fundamentals.

Kyle, Albert S., and Wei Xiong. 2001. "Contagion as a Wealth Effect." *Journal of Finance* 56 (4): 1401 40. A theoretical model that demonstrates the role of investors' wealth constraints in generating return comovements among assets not sharing any common economic fundamentals.

Yuan, Kathy. 2005. "Asymmetric Price Movements and Borrowing Constraints: A Rational Expectations Equilibrium Model of Crises, Contagion, and Confusion." *Journal of Finance* 60 (1): 379 411. A theoretical model that demonstrates the role of investors' liquidity constraints in generating return comovements among assets not sharing any common economic fundamentals.

JON WONGSWAN

■ steel

Steel, one the most basic industrial commodities, has been the subject of numerous international trade disputes and policy interventions throughout the 20th and early 21st centuries. Because steel is a crucial part of many other manufactured products and of construction activity, policymakers often consider it

to be a key sector for a nation's industrial development. Public support for steel industries has also responded to the recognition that jobs at steel plants usually offer higher-than-average wages, or what economists call "labor rents" (Scott and Blecker 1997). However, many economists consider the efforts of governments to protect or subsidize their national steel industries to be examples of inefficient, socially costly interventions that benefit small groups of workers and firms at the expense of consumers (see de Melo and Tarr 1993).

Technology, Costs, and the Location of Production Steel is a metal alloy of iron, carbon, and other minerals that has tremendous strength and ductility and is thus useful for a wide range of purposes from bridges and skyscrapers to motor vehicles and oil pipelines. Steel production developed during the industrial revolution of the 19th century in response to the discovery that pure iron was too brittle to be reliable for railroad tracks, steam engines, and other applications subject to stresses of heat and pressure. Steel is not a single product; the term encompasses a variety of specific types (for example, rust-resistant galvanized products and stainless steel for decorative uses).

Traditionally, basic steel production takes place in a two-step or "integrated" process. In the first step, iron ore is reduced to molten ("pig") iron in a blast furnace, using coke made from coal as fuel; in the second step, the molten iron is transformed into "crude" or "raw" steel in a steel-making furnace. After crude steel is produced, it is then subject to further casting, rolling, coating, and shaping operations to transform it into useful forms, such as sheets, girders, pipes, plates, tin-coated cans, and wire rods.

The progression from the Bessemer and Thomas steel furnaces of the 19th century to the open-hearth furnace and later the basic oxygen furnace in the 20th century is considered one of the classical examples of process-oriented technological change. In the second half of the 20th century, another important innovation was the electric arc furnace, which produces crude steel from "scrap" (recycled steel) in a one-step process with little or no iron ore and lower capital costs. By 2006, electric furnaces accounted for 57

percent of U.S. production and 32 percent of global production.

Originally, steel mills had to be located near sources of their principal raw materials, iron ore and coking coal, due to high transportation costs. However, advances in ocean shipping in the 20th century revolutionized the location of integrated steel production, allowing resource-poor countries such as Japan to produce steel efficiently using imported raw materials, provided that their steel mills were located near seaports. In addition, "minimills" that use the electric furnace technology can be located farther from resource deposits or seaports and closer to downstream users. These technological developments have led to greater geographical dispersion of steel production both internationally and regionally within nations.

It is often popular in the United States to blame losses of steel jobs on import competition. Although imports have adversely affected steel employment in some short-run situations, by far the main causes of decreased U.S. steel employment in the long run have been the dramatic increase in labor productivity in the industry and the slow growth of domestic steel demand (Grossman 1986). In 1955, the U.S. steel industry employed 707,000 people to make 77 million metric tons of shipments, while 50 years later in 2005 the industry needed only 122,000 employees to produce 95 million metric tons. Using a more refined measure that controls for product composition, the U.S. Bureau of Labor Statistics index of labor productivity in steel production shows a four-fold increase between 1950 and 2000.

Historically, steel industries generated large amounts of employment and were bastions of strong labor unions that won high wages and benefits for their members. The enormous shrinkage in U.S. steel employment in the late 20th century led to the problem of high "legacy costs" of pensions and benefits for retired steelworkers at some large U.S. steel corporations in the 1990s; the U.S. government addressed this problem by having the Pension Benefit Guarantee Corporation take over some of those obligations. However, labor costs of currently employed workers are now a relatively small portion of

variable costs in the most modern steel mills where production is so automated that a small number of skilled workers can operate the equipment using computerized controls and many minimill firms are nonunion.

Steel production was originally concentrated in the technologically leading nations, led by Britain and followed by the United States, Germany, and other European nations, in the late 19th and early 20th centuries. As the technology matured during the 20th century, steel production spread to Japan, the Soviet Union, China, and the major developing nations (especially India, Brazil, South Korea, Turkey, Taiwan, and Mexico). As of 1965, the United States and Soviet Union were the two largest producers of crude steel, with 26 percent and 20 percent of total world output, respectively. By 2006, China had become the world's largest producer, accounting for 34 percent of global production; Japan and the United States ranked second and third at 9 percent and 8 percent, respectively.

The Role of Trade Policies Trade and industrial policies have long been key determinants of the location of steel production. In the late 19th and early 20th centuries, the United States, Germany, and other Western countries used tariffs to protect their "infant" steel industries, a policy that was followed later in the 20th century by Japan, India, Brazil, South Korea, and many others. Communist countries such as the former Soviet Union and the pre-reform People's Republic of China created steel industries based on state-owned enterprises managed through central planning. Steel was one of the most favored industries under "import substitution" policies in many developing nations between (roughly) the 1940s and 1970s and subsequently has been a frequent target of export-promotion efforts in many of these same nations.

Curiously, out of the five largest exporters of steel as of 2005, two countries had some of the most advanced steelmaking facilities in the world (Japan and Germany), while two others had some of the most technologically backward facilities (Russia and Ukraine, which were still using open-hearth furnaces for significant portions of their output); China, the third largest exporter, had a mix of older and newer steelmaking plants. As would be expected, Japan and Germany export relatively more specialty and high value-added steels, while Russia and Ukraine export mostly basic, commodity-grade products, and China is in the process of upgrading its exports.

Given that steelmaking technology is highly mobile and the main inputs (both raw materials and machinery) can be imported, some economists argue that there are no inherent (resource-based) "comparative advantages" in steel today and that competitive industries can be created in a wide range of countries. This view has been used (for example, by Howell et al. 1988) to justify government intervention to protect the U.S. industry against "unfair" competition from other countries that subsidize or protect their industries, or that gain artificial advantages through means such as undervalued currencies. Some East Asian countries have been accused of having "structural impediments" to steel trade because their vertically integrated industrial groups (*keiretsu* in Japan, *chaebol* in Korea) allegedly do not buy steel from outside companies. Other economists counter that steel exports are still explained by comparative advantage, because steel is exported by the countries that can make it relatively cheaper than other products, compared with other countries, regardless of the source of that relative cost advantage. In this latter view, efforts to protect domestic steel industries merely make steel more expensive for domestic consumers and downstream industries.

The nature of steel production has made the industry the subject of many international trade disputes. Steel production, especially in integrated mills, is capital intensive and has large economies of scale, which create a tendency toward the existence of excess capacity (except in times of strong demand). As a result, steel producers often have incentives to export as much as possible in order to bolster capacity utilization, which leads to accusations of "dumping" if the products are sold at lower prices in foreign markets. At the same time, import-competing steel producers also have incentives to maintain high rates of capacity utilization and are therefore likely to seek trade protection if they lose customers to imports.

When one adds in the popular view of steel as a "strategic" product for reasons of both economic development and national security, as well as union support for high-wage jobs, one has an industry in which policy intervention has generally been the norm rather than the exception.

Since World War II, multilateral trade liberalization under the General Agreement on Tariffs and Trade and subsequently the World Trade Organization (WTO) has greatly reduced or, for some products, completely eliminated the ordinary tariffs that formerly were used to protect national steel industries. Nevertheless many steel products continue to receive "administered" or "contingent" protection, mainly through the use of anti-dumping and countervailing duties.

The United States has adopted a variety of "trade remedies" for steel producers since imports achieved significant levels in the 1960s, including voluntary export restraints (VERs) in 1969 74, trigger prices (import price floors) in 1978 79, more VERs in 1982 91, and a safeguard tariff in 2002 3. In addition, many U.S. imports of particular steel products from specific countries have been subject to anti-dumping or countervailing duties, especially since the last VERs expired in 1992. Iron and steel mill products (excluding castings) accounted for 44 percent of all U.S. anti-dumping and countervailing duty orders in effect as of 2006. In spite of these trade barriers, however, the United States remained by far the largest net importer of steel as of 2005.

The United States is far from alone in its use of trade protection for its steel industry. The European Union (EU), Canada, and other industrialized countries have also imposed anti-dumping duties on specific steel imports, especially from newly industrializing countries and transition economies, and the latter groups of countries have begun to impose anti-dumping duties of their own.

An important change in recent decades is the increase in "vertical" trade in steel that is, importing intermediate steel products to manufacture finished or semifinished products. For example, "slabs" of crude steel can be imported for flat-rolling into steel sheet, or steel sheet can be imported for the pro-

duction of pipe and tube products. This change may alter the political economy of steel trade policy, since companies that rely heavily on such imports may not support protection of the intermediate products they import even though they may still seek protection for their own finished goods.

Changing Competitive Conditions In roughly the first six decades of the 20th century (and longer in some developing countries), the combination of large-scale economies, high transportation costs, and significant tariff barriers led to the creation of national steel industries that were relatively closed to international trade and dominated by small numbers of large firms. Steel companies were frequently cited as exemplars of firms that practiced collusion, price leadership, administered pricing, cartels, or other forms of oligopolistic behavior. Large oligopolistic steel firms in some countries were accused of being slow to adopt new technologies such as the oxygen furnace in the 1960s.

Since the 1960s, several of the structural changes referred to earlier have led to a more competitive market structure for most countries' steel industries (Barnett and Schorsch 1983; Blecker 1989). Minimills took advantage of lower unit costs to compete effectively with the older, integrated firms. Trade liberalization and reduced transportation costs, along with the diffusion of steelmaking technology to less-developed countries, led to increased international competition. As a result, basic steel products became "commodities" in the sense that their prices were determined by global supply and demand conditions, with little or no oligopoly power for even the largest national firms in most product lines (except where those firms enjoyed continued protection).

In Europe, the steel industry was at the center of efforts to promote regional integration in the post World War II period. The European Coal and Steel Community (ECSC), founded in 1951, combined six major countries in an effort to rationalize their steel industries and make them more competitive while ameliorating the social costs of adjustment. The ECSC was an important forerunner of the European Economic Community and later the EU. As a result of these integration efforts, European steel

firms began to consolidate on a multinational basis somewhat earlier than steel firms in other regions. However, joint ventures and foreign direct investment also increased in other areas of the world steel industry toward the end of the 20th century.

By the early 2000s, the steel industry was beginning to reconsolidate on a global scale. Through mergers and acquisitions, a number of large multinational corporations were formed with steel operations located across Europe, Asia, and the Americas. As of 2007, the world's largest steel company was ArcelorMittal, which had plants in 27 countries, employed 320,000 workers, and produced nearly 10 percent of total world steel output. Headquartered in London and Luxembourg, ArcelorMittal was formed by the merger of two former companies, both of which had grown through previous mergers and acquisitions, and one of which was founded by an Indian-born entrepreneur (Lakshmi Mittal).

In the late 1990s, the global steel market was afflicted by excess supply conditions, as a number of transition economies (led by China, Russia, and Ukraine) increased the world supply of steel at the same time as international demand conditions weakened due to financial crises in East Asia and elsewhere. This resulted in increased exports, falling prices, lower profits, and increased trade tensions. The U.S. safeguard tariffs of 2002 3 were a lagged response to this crisis, adopted after the recession of 2001 and the appreciation of the U.S. dollar further weakened the U.S. industry, and when a new president (George W. Bush) was more willing to support a safeguard remedy than his predecessor (Bill Clinton) (see Blecker 2008). Ultimately, these tariffs were ruled illegal by the WTO in 2003, partly because they were enacted so long after U.S. steel imports peaked in 1998. Liebman (2006) found that the safeguard tariffs, which covered only about one-third of U.S. steel imports, had little effect on domestic prices.

In a stunning reversal, global demand for steel began to outstrip global supply in 2004, leading to sharp price increases that restored industry profitability worldwide for the next few years. This reversal was due partly to an enormous increase in demand by China, where steel consumption nearly tripled be-

tween 2000 and 2006. However, steel producers around the world responded to the higher prices by building new capacity, and as a result the global industry is likely to experience more boom-bust cycles in the future.

In spite of its reputation as an "old" or "smokestack" industrial product, steel remains an important internationally traded commodity. World steel production grew by about 50 percent from 2000 to 2006, and almost 40 percent of global steel production is exported. Therefore, the global steel industry is likely to remain at the center of trade policy disputes for the foreseeable future.

See also anti-dumping; countervailing duties; European Union; export promotion; foreign direct investment and international technology transfer; import substitution industrialization; infant industry argument; joint ventures; mergers and acquisitions; political economy of trade policy; safeguards; shipping; technological progress in open economies; transition economies; World Trade Organization

FURTHER READING

Barnett, Donald F., and Louis Schorsch. 1983. *Steel: Up heaval in a Basic Industry*. Cambridge, MA: Ballinger. A comprehensive account of international and techno logical factors in the evolution of the U.S. and global steel industries.

Blecker, Robert A. 1989. "Markup Pricing, Import Com petition, and the Decline of the American Steel In dustry." *Journal of Post Keynesian Economics* 12 (1): 70 87. An econometric analysis of how import competition reduced oligopoly power in the U.S. steel industry.

. 2008. "U.S. Steel Import Tariffs: The Politics of Global Markets." In *Contemporary Cases in U.S. Foreign Policy: From Terrorism to Trade*, edited by Ralph G. Carter. 3d ed. Washington, DC: Congressional Quar terly Press, 249 80. An examination of the political economy of the U.S. decision to impose safeguard tariffs in 2002 3.

Cockerill, Anthony, and Aubrey Silberston. 1974. *The Steel Industry: International Comparisons of Industrial Struc ture and Performance*. London: Cambridge University Press. A definitive study of scale economies and pro duction costs.

de Melo, Jaime, and David Tarr. 1993. "Industrial Policy in the Presence of Wage Distortions: The Case of the U.S. Auto and Steel Industries." *International Economic Re view* 34 (4): 833 51. A study of the welfare effects of steel protection using a computable general equilibrium model.

Grossman, Gene M. 1986. "Imports as a Cause of Injury: The Case of the U.S. Steel Industry." *Journal of Inter national Economics* 20 (3/4): 201 23. An econometric study that pioneered the use of a reduced form model to estimate the causes of declining steel employment.

Howell, Thomas R., William A. Noellert, Jesse G. Kreier, and Alan Wm. Wolff. 1988. *Steel and the State: Gov ernment Intervention and the Steel Crisis.* Boulder, CO: Westview. A study of government efforts to promote national steel industries in numerous countries around the world.

International Iron and Steel Institute (IISI). 2007. *World Steel in Figures 2007.* Brussels: IISI. www.worldsteel.org. The standard statistical reference on the global steel in dustry (issued annually).

Liebman, Benjamin H. 2006. "Safeguards, China, and the Price of Steel." *Review of World Economics* 142 (2): 354 73. An econometric study of the effects of safeguard tariffs and other factors on steel prices in the early 2000s.

Scott, Robert E., and Robert A. Blecker. 1997. "Labour Rents, Adjustment Costs, and the Cost of U.S. Steel Restraints in the 1980s." *International Review of Applied Economic*s 11 (3): 399 419. Estimates of the social costs and benefits of alternative U.S. steel trade policies.

ROBERT A. BLECKER

■ sterilization

Sterilization is the process by which monetary authorities ensure that foreign exchange interventions do not affect the domestic monetary base, which is one component of the overall money supply. Many governments try to influence the value of their currency on the foreign exchange market by selling or purchasing domestic currency in exchange for a foreign currency. If the monetary authority sells domestic currency that was not previously in circulation, the intervention will expand the domestic money base. Likewise, if the monetary authority purchases more of its domestic currency and takes the receipts out of circulation, the intervention will have a contractionary effect on the domestic money base.

Sterilized intervention operations involve domestic asset transactions that restore the monetary base to its original size. For example, a nonsterilized sale of foreign currency on the open market would result in a reduction in the central bank's net foreign assets (NFA) and a contraction of the domestic monetary base (MB). This operation can be sterilized, or neutralized, by an offsetting purchase of domestic currency that increases the central bank's net domestic assets (NDA) and returns the monetary base to its original level.

In theory the process of sterilization is quite straightforward, but in practice it may be difficult for the monetary authority to fully offset the effects of a change in net foreign assets. In countries with less-developed financial markets the ability to sterilize may be constrained by the size and depth of the domestic bond market. Additionally, monetary authorities may not be able to sterilize intervention operations in fixed exchange rate systems with some degree of capital mobility (Obstfeld 1982). For example, sales of domestic-currency assets will attract a capital inflow, forcing the authorities to buy more foreign assets in order to maintain the fixed value of the currency, thereby offsetting any attempt to sterilize the original open-market asset sale.

Costs of Sterilization Sterilization may also come at a fiscal cost. Governments attempting to lower or maintain the value of their domestic currency in the face of market pressure for a domestic currency appreciation will generally be purchasing relatively low-yield foreign assets while selling relatively high-yield domestic assets. The fiscal burden of sterilization will depend on the interest differential between the domestic and foreign assets. Further, the international accumulation that results from these sorts of sterilization operations will expose the government to foreign exchange risk. If the domestic currency eventually appreciates relative to the foreign currency denominations in a country's reserves, the country will experience a capital loss. On the other

hand, governments attempting to prevent a depreciation of their domestic currency will generally be selling foreign assets and purchasing domestic assets. The constraint in this case will be the size of the country's foreign reserves.

The Efficacy of Sterilized Intervention Operations In most monetary and asset-pricing models of exchange rate determination, nonsterilized intervention will affect the exchange rate in proportion to the change in the relative supplies of domestic and foreign money, just as any other form of monetary policy does. The effectiveness of sterilized intervention operations in standard models depends on two additional assumptions: that domestic and foreign bonds are outside assets (i.e., the public considers these bonds as net wealth) and that they are imperfect substitutes (meaning that the currency denomination of the bonds matters to investors). Sterilized intervention can also influence the exchange rate in models where the government is assumed to have more information about relevant economic fundamentals (such as future money and income differentials) than the market and can credibly convey that information using intervention operations.

Governments generally finance their spending by raising taxes or borrowing by issuing bonds. If they issue bonds, the public has more money to spend. Further, if the public ignores the fact that taxes will need to be raised in the future to pay off the bonds, these bonds can be considered "outside assets" and are additions to net wealth. On the other hand, if the public recognizes that they (or future generations) will have to pay higher taxes in the future and therefore save the extra money in order to pay the future tax, bonds are "inside assets" and cannot be considered net wealth. The extra saving by the public would exactly offset the extra spending by the government, so overall demand would remain unchanged. This view of the implications of bond financing is termed Ricardian equivalence. Sterilized intervention operations in such a world are simply swaps in the currency composition of inside assets, and these should have no effect on the foreign exchange market equilibrium.

Even if it is granted that government bonds are outside assets, sterilized intervention will have no effect on the exchange rate if domestic and foreign bonds are perfect substitutes. If investors are completely indifferent between holding domestic and foreign bonds, then changes in their relative supply should have no effect. If bonds are not perfect substitutes, however, even if they are close substitutes, then changes in bond supplies matter, so changes in their relative supply can influence the exchange rate through the portfolio-balance channel.

In portfolio-balance models of exchange rate determination, investors diversify their holdings among domestic and foreign bonds as a function of both expected returns and the variance in returns. By changing their relative supply, sterilized intervention operations alter the risk characteristics of foreign and domestic bonds in the market portfolio, and thus alter the equilibrium exchange rate. For example, a sterilized sale of domestic-currency-denominated bonds may increase their relative riskiness because investors will be more vulnerable to unexpected changes in the value of the domestic currency. Investors will require a higher expected return on domestic bonds to hold willingly the larger outstanding stock, leading to a depreciation of the domestic currency.

Finally, even for those who hold either to the Ricardian equivalence or to the assumption that foreign and domestic bonds are perfect substitutes, sterilized intervention can have an effect on exchange rates if it provides the market relevant information that was previously not known or not fully incorporated in the current exchange rate. The information channel for sterilized intervention is controversial. It relies on the existence of an asymmetry between what is known by the government and what is known by market participants. In order for sterilized intervention operations to influence exchange rates via the information channel, the government must both have inside information and have the incentive to reveal the information truthfully by way of their operations in the foreign exchange market. Indeed, it has been suggested that sterilized intervention may be used by governments to " buy credibil-

ity" for their future policy intentions (Mussa 1981). If market participants believe the signals provided by sterilized intervention, they will influence exchange rates by betting with the operation.

The information channel for sterilized intervention need not exclusively serve to convey future policy intentions. For instance, intervention helps to convey information by the monetary authorities to the market in circumstances when such information might not be made directly available for security or other reasons (Friedman 1953). Intervention signals may also alter the market's expectations, especially when market participants are heterogeneous and there are signs of a bubble developing (Kenen 1987). As long as the information signaled through sterilized intervention policy is relevant and credible, it can potentially influence the exchange rate. If the information revealed involves the monetary authority's own future policy intentions, however, then sterilized intervention should not be considered an additional independent policy tool. The sterilized intervention operation may alter the timing or magnitude of the impact of monetary or fiscal policy on the exchange rate, but its effectiveness is not independent of those policies.

Empirical Evidence Is there empirical evidence that sterilized intervention operations affect exchange rates? In 1982 the Group of Seven (G7) economic summit at Versailles commissioned a comprehensive study of intervention policy in order to answer the question. The G7 working group report, completed in 1983, draws no explicit conclusions but suggests that the effects of sterilized interventions on the exchange rate were small and transitory at most over the period 1973 81 (Jurgensen 1983). Subsequent studies of G3 intervention policy suggest that more recent operations may have been more effective (Dominguez 1990, 2003, 2006; Dominguez and Frankel 1993a, 1993b; Sarno and Taylor 2001).

An indirect test of the efficacy of sterilized intervention involves examining whether the assumptions underlying the portfolio-balance channel are satisfied in the data. In particular, one such test examines whether foreign and domestic bonds are imperfect substitutes in investors' portfolios. If investors are indifferent between holding domestic assets and foreign assets, then once we take into account both the current and expected exchange rate, there should be no return differential (or risk premium) between the two. This hypothesis is commonly referred to in the literature as the uncovered interest parity condition. Most empirical tests of uncovered interest parity find that foreign and domestic bonds are not perfect substitutes.

The failure of uncovered interest parity is a necessary but not a sufficient condition for sterilized intervention to affect the exchange rate through the portfolio-balance channel. There must be a stable relationship between government debt supplies and the return differential between domestic and foreign bonds, and empirical studies have had mixed success relating the two. Studies that include government debt and other outside assets, which usually dwarf foreign exchange intervention in magnitude, in the definition of government debt supplies reject the hypothesis that the two are related. Empirical studies using daily (and intradaily) intervention data, which are able to focus exclusively on short-term changes in asset supplies through foreign exchange intervention, generally find evidence to support the hypothesis that sterilized intervention operations systematically influence return differentials (Dominguez 1990, 2003, 2006; Dominguez and Frankel 1993a, 1993b).

In order for the information channel to be operative, sterilized intervention operations must be observed by market participants. If only for this reason, it is puzzling that more governments do not disclose data on their intervention operations. Comparisons of actual intervention data with newswire reports of intervention suggest that G3 operations are generally reported in the financial media. Empirical tests that distinguish interventions that are reported (and therefore are capable of serving as signals) from those that remain secret generally find that it is mainly reported interventions that significantly influence exchange rates, providing indirect evidence for the signaling channel. The literature, however, has been less successful at finding a systematic link between reported interventions and future fundamentals (the information that is supposedly being conveyed by the

government), making it difficult to find direct evidence for signaling (Sarno and Taylor 2001).

Theory suggests that sterilized intervention operations can potentially provide governments an additional policy tool with which to attain internal and external balance. Practice suggests that governments in both fixed and flexible exchange rate systems have frequently resorted to sterilized intervention policy. Empirical studies of the efficacy of these operations suggest that intervention in developed countries has often been successful, though whether sterilized intervention can serve as a fully independent policy tool remains controversial. The efficacy of sterilized intervention policies in developing countries has been less widely studied, in large part because governments have been reluctant to provide data on their operations. Developing countries are also not always able to fully sterilize their operations due to their illiquid domestic bond markets and the potentially high fiscal costs of foreign reserve accumulation.

See also asymmetric information; balance of payments; equilibrium exchange rate; exchange rate regimes; exchange rate volatility; foreign exchange intervention; interest parity conditions; international reserves; money supply; purchasing power parity; real exchange rate; twin deficits

FURTHER READING

Dominguez, Kathryn M. 1990. "Market Responses to Coordinated Central Bank Intervention." *Carnegie Rochester Series on Public Policy* 32: 121 64. One of the first empirical studies of the efficacy of the intervention signaling channel.

. 2003. "Foreign Exchange Intervention: Did it Work in the 1990s?" In *Dollar Overvaluation and the World Economy*, edited by Fred Bergsten and John Williamson. Special Report 15. Institute for International Economics. Washington, DC: IIE, 217 45. Updates the empirical work in Dominguez and Frankel (1993b) through 2002.

. 2006. "When Do Central Bank Interventions Influence Intra Daily and Longer Term Exchange Rate Movements?" *Journal of International Money and Finance* 25: 1051 71. Examines both market micro structure and longer term effects of G3 intervention operations since 1989.

Dominguez, Kathryn M., and Jeffrey Frankel. 1993a. "Does Foreign Exchange Intervention Matter? The Portfolio Effect." *American Economic Review* 83, 1356 69. The first study to challenge the conventional wisdom exposed in the Jurgensen report that the portfolio balance effect is economically unimportant.

. 1993b. *Does Foreign Exchange Intervention Work?* Washington, DC: Institute for International Economics. A comprehensive survey of theory and experience regarding exchange market intervention.

Friedman, Milton. 1953. "The Case for Flexible Exchange Rates." In *Essays in Positive Economics*, edited by Milton Friedman. Chicago: University of Chicago Press, 157 203. The classic reference for arguments on why flexible exchange rate regimes are better than fixed regimes.

Jurgensen, Philippe. 1983. "Report of the Working Group on Exchange Market Intervention [Jurgensen Report]." Washington, DC: U.S. Treasury Department. Suggests that the influence of intervention on exchange rates through the portfolio balance channel is economically unimportant.

Kenen, Peter. 1987. "Exchange Rate Management: What Role for Intervention?" *American Economic Review* 77 (2): 194 99. Describes three ways (by brute force, by altering asset market equilibrium, or by altering expectations) that intervention operations can be used by governments to manage exchange rates.

Mussa, Michael. 1981. "The Role of Official Intervention." Group of Thirty Occasional Papers No. 6. New York: Group of Thirty. Provides one of the first informal expositions of the signaling hypothesis.

Obstfeld, Maurice. 1982. "Can We Sterilize? Theory and Evidence." *American Economic Review* 72 (2): 45 50. Reviews the empirical and theoretical scope of sterilization in the foreign exchange market.

Sarno, Lucio, and Mark P. Taylor. 2001. "Official Intervention in the Foreign Exchange Market: Is It Effective and, If So, How Does It Work?" *Journal of Economic Literature* 39: 839 68. An excellent survey of recent research on exchange rate intervention policy.

KATHRYN M. E. DOMINGUEZ

■ **stock versus flow of foreign direct investment**

See foreign direct investment (FDI)

■ **Stolper-Samuelson theorem**

See Heckscher-Ohlin model

■ **strategic trade theory**

See New Trade Theory

■ **structural adjustment**

Since the 1980s, the expression *structural adjustment* (SA) has been used to denote programs of policy reforms in developing countries undertaken with financial support from the World Bank. Structural adjustment programs (SAPs) would normally consist of two components: macroeconomic stabilization and microeconomic, supply-side reforms. The World Bank *Operational Manual* defines structural adjustment lending (SAL) as "non-project lending to support programs of policy and institutional change necessary to modify the structure of an economy so that it can maintain both its growth rate and the viability of its balance of payments in the medium term" (*Operational Manual*, Statement No. 3.58, Annex II, November 1982).

Through SAL the World Bank financed more than 650 reform programs. SAL represented about one-quarter of its total commitments from the mid-1980s to the late 1990s, when it increased to more than 50 percent as a response to the financial crises in East Asia and remained at about one-third thereafter. Sub-Saharan Africa accounts for 34 percent of adjustment operations and 16 percent of commitments. The Latin American and Caribbean region accounts for 24 percent of adjustment operations and, with 34 percent of lending commitments, is the largest recipient of SAL, followed by Europe and Central Asia with 24 percent; East Asia and the Pacific received 15 percent of SAL, South Asia 6 percent, and the Middle East and North Africa 5 per-

cent. SA shaped the strategy of the World Bank for nearly two decades and dominated policymaking in the developing world. During this period, almost all major programs of policy reforms were supported by SALs from the World Bank.

The bank started SAL in 1980 following serious balance of payments difficulties experienced in some of its member countries as a result of the second oil price shock in 1979. Many of these countries also had fiscal deficits and often high inflation, and the root of the external deficits was identified as excessive aggregate demand. In these circumstances, adjustment requires macroeconomic stabilization in order to reduce domestic demand to a level that is consistent with the level of external resources available to the country.

Since restrictive monetary and fiscal policies that cut demand (such as tax increases, spending cuts, and interest rate increases) also cause unemployment, they are normally accompanied by a currency devaluation that, by lowering the relative price of domestic goods with respect to foreign goods, leads to an increase in the foreign demand for domestic goods and an increase in the supply of exportable and import-substitute goods, and minimizes the employment costs of stabilization. Measures to restrain wages complement exchange rate policy to ensure that the competitive effect of nominal devaluation is not offset by wage inflation.

Macroeconomic stabilization alone may not give rise to faster growth in the future. In order to achieve this, a second type of adjustment is required, namely SA, which consists of microeconomic and institutional reforms designed to strengthen the supply response of the economy by letting the market determine the efficient allocation of resources. A central aspect of SA was market liberalization and deregulation along several dimensions, potentially including removal of price controls, deregulation of the domestic goods market, liberalization of the trade regime, deregulation of the domestic financial market, liberalization of the capital market   especially the removal of barriers to foreign direct investment and deregulation of the labor market. A second component involved reform of public sector

management, including fiscal reform, restructuring and privatization of state-owned enterprises, and restructuring of government spending. A final component consisted of reforms to create and strengthen institutions that would support stabilization and SA, such as institutions for banking supervision and the enforcement of property rights, an independent judiciary, and an independent central bank. Many economists believed that these microeconomic reforms would shift the economy to a sustainable path of higher economic growth.

The Rise of Structural Adjustment Lending Several factors contributed to the establishment of SAL. One was the nature of the crisis that was affecting developing countries. Severe balance of payments deficits required faster and larger disbursements than were possible under the standard World Bank approach of project-based lending, in which disbursements are typically slow in the first two years. By contrast, SALs typically disburse within 18 months.

A second factor was the realization within the World Bank that project success was crucially dependent on the state of the broader economic environment. This implied, first, that the World Bank could no longer encourage the pursuit of economic growth without regard for macroeconomic stability, which thus became a priority over growth and was seen as the necessary foundation for sustainable growth. A second implication is that, as the development strategies of many developing countries were now being seen as highly distorted, the World Bank had to turn its attention to removing microeconomic distortions.

SA embodies the drastically new viewpoint that development is hampered not so much by capital shortage as by domestic economic policies that impede the operation of market forces. This new vision was influenced by neoliberal political economy theories that centered on the capture of the state by powerful pressure groups. The state came to be seen as the primary cause of the distortion of incentives, leading to misallocation of resources and economic failure. Free markets work well, and even where there are externalities and market failures, the state does not have the willingness or indeed the capacity to lead

to a better outcome than the market would. This is a complete overturning of the view prevailing until the 1970s, according to which state intervention was necessary in order to correct distortions and facilitate structural transformation in developing countries, where free markets do not work well. In this older view, the state was a benevolent agent for development. By contrast, SA through liberalization, deregulation, and privatization aims to minimize state intervention in the economy and liberalize prices so that they can reach equilibrium levels.

Abandoning structuralist doctrines, which emphasize the existence of special characteristics in developing countries' economic and social structures and were the mainstream in development thinking in the 1960s and 1970s, the World Bank through SA presents a model for all countries to follow, irrespective of their development stage. Indeed SAPs have been characterized by a remarkable degree of uniformity.

With the outbreak of the debt crisis in the 1980s, the main objective of SAL was to prevent an international banking crisis by allowing debt restructuring and the continuation (or the resumption) of interest payments. The objective of poverty reduction one of the bank's priorities since the late 1960s was downgraded. Partly this was also a reflection of the negative assessment within the World Bank of the strategy pursued until then of attacking poverty directly. In addition, this critical assessment led to the view that the best way to reduce poverty is through the promotion of economic growth (the so-called trickle-down doctrine). Thus the policies of SA, which are designed to stimulate higher sustainable growth, may be justified in terms of greater equality and poverty reduction. For example, minimizing state intervention leads to higher growth and, through the trickle-down effect, to lower poverty. Moreover since poverty is particularly acute in rural areas, reducing state intervention which typically favored industry over agriculture lowers poverty and improves income distribution.

Empirical Assessment of the Effects of SAPs The empirical literature in this area is extensive and

its findings are diverse, due to the variety of approaches used in the analyses and the methodological difficulties inherent in the evaluation of SAPs. Any short summary of the evaluation results is therefore inevitably partial. Generalizing and hence partly ignoring the diversity of individual countries' experiences, SAPs appear to have had their stronger positive effects on countries' external accounts (with increases in the rates of growth of exports and improvements in the balance of payments, though such positive results have also been found to be often temporary) rather than on the domestic economic performance (as typically investment fell and economic growth contracted or stagnated).

Indeed, growth in countries that have followed SA policies   especially in Latin America and Africa   was frequently lower than in the 1960s and 1970s, when these countries were following the highly distortionary policies that SA was meant to correct. Both in Latin America and Africa there have been instances when SA seemed to lead to economic success, but such cases were both isolated and short lived. Moreover the 1990s saw the countries that liberalized the domestic financial system and opened the capital account being hit by severe financial crises. The financial fragility of liberalized markets also seemed to make it possible for financial crises to propagate by contagion. In contrast to the experience of the countries that followed SA reforms, the countries that have managed to achieve sustained high growth are those that have significantly deviated from SA, for example by maintaining selective protection of infant industries, by maintaining controls on capital flows, or by liberalizing trade and capital movements at a slow pace.

SAPs also appear to have had negative effects on income equality and the general standards of living, as measured by health, education, and nutrition indicators. Cross-country analyses on poverty and income distribution do not provide a unified picture of worldwide trends (the different results are due to problems with the reliability and comparability of data, different country and time period samples used in the analyses, and different methodological approaches). There is evidence, however, of increases

since the 1980s both in cross-country and within-country inequality as well as in poverty (once China is excluded from the sample). Such trends are attributed to many different factors but some of these factors are related to the adoption of SA policies, such as overly rapid trade and financial liberalizations, rapid and poorly designed privatization programs, and the erosion of labor institutions and of the state's redistributive role. A significant amount of empirical evidence also calls into question the effectiveness of the trickle-down mechanism: increases in average income do not appear to raise the income of the poor for prolonged periods.

What policy implications can be drawn from such empirical findings is a highly controversial matter for two related reasons: first, these findings may be the results of flawed evaluation methodologies, and second, the empirical results may be interpreted and explained in many different ways. Let us consider, for example, the findings concerning the effects of SAPs on economic growth. The expectation is that SAPs would result first in an economic contraction in the short run (as a result of the macroeconomic stabilization), which would then be followed by a recovery (led by the supply-side reforms) that shifts the economy onto a path of sustainable growth. It is clear therefore that SAPs cannot be criticized for causing a contraction, since this is inevitable if the cause of the crisis is an excess of demand over supply. Thus the relevant question to investigate is not whether SAPs caused a contraction but whether the contraction was deeper and/or lasted longer than necessary. Actual program effects should then be compared with what the economy could have done without an SAP (such a hypothetical outcome is called a counterfactual in the literature). Methodologically, the difficulty is that the counterfactual is not observable and can only be estimated.

The different evaluation methodologies can be seen as different approaches for the estimation of the counterfactual. Intuitively, such estimation is based on the economic situation in program countries during the period preceding the start of the SAP or on the economic situation in countries that did not enter a SAP. More sophisticated methodologies would

account for the fact that program and nonprogram countries may be systematically different. For example, program countries are more likely to have been hit by a negative shock in the preprogram period and, consequently, there may be systematic differences in the policies pursued by the two groups of countries during the adjustment period. Alternately, program and nonprogram countries may differ in the extent of their relative policymakers' commitment to good policies. None of these different approaches is fully satisfactory as they all have important shortcomings. Moreover since in any selected sample period not all of the observed growth rates may be sustainable in the long run, ascertaining the true influence of policies on economic growth becomes very difficult. By implication, whatever empirical evidence is found cannot be clear-cut, because it is not possible to know whether it represents the true program effects or is the result of a wrong estimation of the counterfactual.

Inevitably, however, the effects of SAPs have raised concerns about poor economic performance, growing inequality, and increasing social problems, as a result of which the development agenda has been significantly broadened, with greater attention being given to the role of institutions and institutional reforms (particularly with respect to the quality of governance) and with poverty reduction becoming a top priority in World Bank operations. In short, in accordance with much of the empirical evidence, there is consensus   both within the World Bank and among its critics   that SAPs have produced disappointing results. This empirical fact has received different explanations. World Bank critics would argue that SA policies are fundamentally flawed and inappropriate to developing countries' circumstances. Others, including the World Bank, would instead point to the lack of government commitment to policy reform and the ineffectiveness of conditionality (i.e., the set of policy reforms that are required for loan disbursement) as policy leverage. The implication of this view is that policy reform programs can be made more effective by allocating financial resources more selectively to countries with good policies and good institutions.

The End of Structural Adjustment As a response to the perceived shortcomings of SAL, the World Bank formally announced its replacement with Development Policy Lending in August 2004. The latter sees reform ownership   which could be defined as the commitment by a recipient country to undertake reforms independently of the incentives provided by lenders   as essential for the success of reform programs. By implication, the promotion of ownership breaks with the idea, typical of the SA era, that there are universal policy prescriptions. Another major difference from SA concerns the characteristics of conditionality: under Development Policy Lending conditionality must be streamlined, so that the conditions focus only on the reforms that are regarded as essential for program success. In addition, conditionality is ex ante in that loan disbursement is made conditional on a country implementing key reforms, rather than promising to do so in the future, as was the case under SA. This new approach clearly implies much greater selectivity in the allocation of World Bank's financial resources. Criticisms center on the feasibility of the principle of selectivity (particularly its lack of credibility, since withholding finance from a country with weak policies and institutions would deteriorate that country's economic and social situation further) and the conflict between ownership and selectivity (since ownership must be more than the freedom to accept the bank's preferred policies).

See also aid, international; aid, international, and political economy; International Monetary Fund (IMF); International Monetary Fund conditionality; trade-related capacity building; Washington consensus

FURTHER READING

Corbo, Vittorio, and Stanley Fischer. 1995. "Structural Adjustment, Stabilization, and Policy Reform: Domestic and International Finance." In *Handbook of Development Economics*, vol. 3B, edited by Jere Behrman and T. N. Srinivasan. Amsterdam: North Holland, 2845   2924. Analytical frameworks supporting the policies of SAPs.

Cornia, Giovanni A., Richard Jolly, and Frances Stewart. 1987. *Adjustment with a Human Face: Protecting the*

Vulnerable and Promoting Growth. Oxford: Oxford University Press. Widely cited critical evaluation, fo cusing on the social impact of structural adjustment.

Fine, Ben, Costas Lapavitsas, and Jonathan Pincus, eds. 2001. *Neither Washington nor Post Washington Con sensus: Challenging Development Policy in the Twenty First Century.* London: Routledge. Critical analyses of the neoliberal approach to policy reform.

Jayarajah, Carl, and William Branson. 1995. *Structural and Sectoral Adjustment: World Bank Experience, 1980 92.* Washington DC: World Bank. Large scale World Bank evaluation.

Koeberle, Stefan, Harold Bedoya, Peter Silarszky, and Gero Verheyen, eds. 2005. *Conditionality Revisited: Concepts, Experiences, and Lessons.* Washington DC: World Bank. Review of the role of conditionality in aid and policy lending.

Mosley, Paul, Jane Harrigan, and John Toye. 1991. *Aid and Power: The World Bank and Policy Based Lending.* London: Routledge. Widely cited critical evaluation, focusing on the relationship between the bank and ad justment lending recipients.

Paloni, Alberto, and Maurizio Zanardi, eds. 2006. *The IMF, World Bank, and Policy Reform.* London: Routledge. Looks at the changes in the intervention strategies of the international financial institutions.

Stiglitz, Joseph. 1998. "Towards a New Paradigm for De velopment: Strategies, Policies, and Processes." Prebisch Lecture at the United Nations Conference on Trade and Development, Geneva, October 19. Criticism of the Washington consensus and proposal of a "post Washington consensus" by the then chief economist at the World Bank.

Williamson, John. 1990. "What Washington Means by Policy Reform." In *Latin American Adjustment: How Much Has Happened?,* edited by John Williamson. Washington, DC: Institute for International Econom ics, 5 20. Lays out the main points of the Washington consensus, an expression used to denote SA reforms.

World Bank. 1998. *Assessing Aid: What Works, What Doesn't, and Why?* New York: Oxford University Press. Final report on a multiyear research program on aid effectiveness.

ALBERTO PALONI

■ **subsidiary**

See multinational enterprises

■ **subsidies and financial incentives to foreign direct investment**

When considering the efforts of governments to attract foreign direct investment (FDI) through the use of subsidies and financial incentives, the game is not new, just the arena in which the game is played. In general, governments influence particular activities and pursue various objectives using instruments including taxes, public spending, regulations, subsidized credit, contingent liabilities, and others. In past decades, when the countries' economies were relatively closed and there was limited trade and scarce capital movements, the policy instruments were used mainly to influence the domestic allocation of resources. Especially in the 1950s and 1960s, governments, particularly those of developing countries, provided generous incentives to domestic enterprises to increase investment in the hope of stimulating higher economic growth.

In countries with federal structures of government, such as the United States, Brazil, Argentina, Canada, and some others, the subnational governments (states, provinces, regions) have often used their own policy instruments to attract economic activities to their area from the rest of the country. In some of these countries there has been intense competition among the states, often pursued through the use of tax incentives and other instruments to attract companies and investments to particular states. This competition has led to distortions in the allocation of resources and has generated a large literature on competition among the states (see, among others, Tanzi 1995, chapter 3; Kenyon and Kincaid 1991).

In today's globalizing world, countries have become more open to international trade flows and capital movements, and capital has become very mobile and highly sensitive to differences in rates of return earned from investing in different locations. Countries have recognized the possibility of attracting capital from the rest of the world through their policies. World capital has become a kind of common

pool from which countries can extract a larger share of investment by pursuing particular policies. Given that most countries are a small part of the world economy, well-designed subsidies and financial incentives to FDI can, at least in theory, attract large amounts of resources relative to country size. These productive activities could create jobs for unemployed workers, or better jobs for underemployed ones, as well as faster growth for countries successful in attracting capital from the global pool. Malaysia, other countries from Southeast Asia, Ireland, and several transition economies from Eastern Europe have been particularly successful in attracting large amounts of FDI in part through the use of incentives directed to FDI. Other countries, including Latin American, African, and especially Caribbean countries have been much less successful in these attempts.

When a country that provides these incentives has many unemployed workers or many workers in unproductive activities that could find employment in more productive activities, the potential benefits for the country from the provision of incentives to FDI can be potentially high. Foreign capital could then be combined with domestic, previously unemployed or underemployed workers to generate new output. Furthermore FDI is often accompanied by new technologies and new managerial techniques that can be copied by domestic firms thus creating beneficial externalities and further growth. This inflow of new knowledge and technologies can, thus, create a theoretical justification for the provision of incentives (see Harrison and Aitken 1999). Even if the foreign enterprises did not pay taxes to the host country on their profits, the country might benefit because the workers would pay taxes on their wages and on their purchases and the domestic firms would become more productive. For these reasons, developing countries are understandably tempted to provide subsidies and financial incentives to FDI.

Types of Incentives There are several kinds of incentives that governments can give foreign investors to induce them to invest in their countries. First, they can provide incentives in kind, especially at the initial stage, to make a particular location more attractive than others to the investors. For example,

land or other real property can be made available free or at low cost and zoning restrictions can be relaxed to allow the foreign investors to build and operate industrial plants in particular areas. Alternatively countries can offer to build needed infrastructures, including roads, railroads, harbors, and so on. These incentives can be particularly valuable in attracting large industrial investments especially to areas that are still underdeveloped. The investments in infrastructures, although costly to the countries, can stimulate other domestic activities and thus create multiplier effects accompanying the foreign investments. In some cases the countries simply accelerate the building of infrastructures that would have been done at some later time.

Second, the foreign enterprises can be exempted from existing regulations, including labor regulations or environmental restrictions. Third, the countries could provide credit at subsidized rates, or facilitate borrowing from domestic sources. This approach would reduce the net capital inflow and the benefit from the foreign investment, however. Finally, the countries could provide incentives vis-à-vis the payment of taxes. These tax incentives are common and are the ones that attract most attention.

The tax incentives may be of various types. Their value to the foreign investors depends on several considerations but the most important is, probably, the level of effective tax rates to which they would be subjected in the absence of the tax incentives. This would depend on the way the taxes paid to the host country (the country where the investment is made) would be treated by the home country (the country of residence of the investor). If the tax incentive results in a reduction in tax payment to the host country but in a higher payment to the home country, the value of the incentive to the investor would be reduced or even eliminated.

The tax incentives can be divided into those against direct taxes and those against indirect taxes (see Zee, Stotsky, and Ley 2002). Among the incentives against direct taxes, the most important are (1) tax holidays against corporate income taxes (CIT); (2) reductions in CIT rates; (3) investment allowances; (4) investment tax credits; (5) accelerated de-

preciation; and (6) investment subsidies. Among the incentives against indirect taxes, the most important are tariff and value-added tax exemptions and incentives connected with export processing zones.

Tax holidays are especially common in developing countries, but much less so in advanced countries. They generally exempt foreign enterprises from the payment of CITs for a given number of years after the investment is made. Thus they are more attractive for those enterprises that expect quick and large profits. Tax holidays have some good and some less good features. Among the good features (for the enterprises) is the fact that during the period of the tax holiday the enterprises not only do not have to pay taxes but they do not need, generally, to comply with all the complexities of tax laws and regulations that can be quite burdensome and costly, especially when dealing with them for the first time in a foreign country. Thus the investors save in terms of both tax liability and tax compliance. Another good feature, from the point of view of the country, is that tax holidays tend to be neutral vis-à-vis the use of capital and labor. They do not distort the allocation of resources in this sense.

There are, however, some less good features of tax holidays. While tax holidays do not distort the choice between labor and capital, they do distort the choice between projects with short maturities and those with long maturities, favoring the former so that the profits earned by the enterprises can occur during the period when the tax holiday is in effect. Second, in practice, the tax holiday can often be extended (at times almost indefinitely) by making additional and often minor investments that allow the enterprise to claim an extension of the tax holiday. For example, in Jamaica and most Caribbean countries, enterprises that invest in hotels and that benefit from tax holidays can renew the tax holiday period by periodically adding a few hotel rooms. Another worrisome feature is that because the enterprises do not need to maintain accounting for tax purposes, there may not be any estimates of the revenue that the country loses due to the tax holidays. Thus it becomes difficult to make a cost-benefit evaluation of the incentives. Finally, in many cases, the enterprises would have

made the investment even without the tax incentives, so that the country forgoes revenue needlessly.

While tax holidays exempt profits completely, *preferential CIT rates* tax their profits at lower rates than for ordinary enterprises. Thus they imply less revenue loss to the governments than tax holidays. They also require that the enterprises keep accounts so that the revenue forgone by the government due to the incentives can be estimated. The difference in tax rates between the enterprises that do not benefit from the incentives and those that do gives a measure of the revenue cost of the incentive under the assumption that the investment would have taken place even without the incentive.

Investment allowances allow the enterprise the immediate expensing of parts of the initial investment costs in addition to the normal allowable depreciation. *Investment tax credits* allow the deduction of parts of the investment costs from the CIT payment. In particular circumstances these two incentives tend to be the same. They are important for enterprises already in operation and that already have profits. *Investment subsidies* involve direct cash payments by the government to an enterprise for parts of an authorized investment. Thus they are a claim on the budgetary revenue of often cash-strapped governments. Finally, economists generally favor incentives given in the form of *accelerated depreciation*. These incentives have fewer disadvantages and lower budgetary costs than other incentives (see Tanzi and Zee 2000).

Indirect tax incentives can be costly to a country in terms of forgone revenue, because indirect taxes are often a large share of the country's gross domestic product (GDP). These incentives imply exonerations from import duties and from other indirect taxes, such as value-added taxes, excises, and other. These incentives lend themselves to abuses because the goods imported, or acquired duty free for the direct use of the enterprises, often find their way to other uses in the domestic market (for a discussion of the potential abuses of these tax incentives, see Zee et al. 2002).

Unfortunately, information on the use of incentives is limited, country specific, and often contained

in confidential reports. Thus it is difficult, if not impossible, to assemble statistical information on the incentives just mentioned for groups of countries. As a consequence, the discussion that follows will not have the desirable backing of statistical information.

Before moving to the next section, it may be useful to mention a worrisome aspect of tax incentives. At times their features are specified precisely and in detail in the laws of the countries, so they have the backing of rule of law. At other times, they are given largely at the discretion of particular offices or officers. The latter may grant them to "essential" activities or "necessary" investments without a clear specification of what these terms mean. Often, the offices that make the decisions are not in the ministry of finance but in spending ministries (tourism, agriculture, industry, etc.) that are not responsible for the outcome of the public finances of the countries but represent sectoral interests. As a consequence, they tend to be generous at times, too generous with the public money.

Furthermore the discretion given to some bureaucrats in making decisions worth a lot of money to particular enterprises makes this an area particularly exposed to corruption. Foreign investors at times find poorly paid, easy to bribe civil servants to get decisions that to them are worth millions of dollars but that may not be particularly beneficial to the countries or decisions that the investors would have made in any case. For this reason, and apart from the intrinsic value of the incentives to a country, the rules that guide the granting of the incentives should be specified precisely in the legislation and should leave as little discretion as possible to civil servants and elected officials. Furthermore, they should be controlled at the highest level by the ministry of finance and possibly by some accountability office.

The Impact of Incentives on FDI Views about the merit of incentives and their impact on FDI diverge widely between those held by most economists who have studied the issues and those held by policymakers. Generally economists have concluded that incentives are often worthless or at least that they do not justify their costs. However, incentives remain popular with policymakers and, of course, with those who benefit from them, namely, the foreign investors. The latter often exert pressure to get the most generous incentives that the host country could provide.

Surveys of foreign investors that ask them the reasons why they have invested in a given country almost never list incentives at the top or near the top of the motivations for the choice of location. The responses mention reasons such as (1) the existence of domestic demand for the product or service produced; (2) the vicinity or the access to other markets; (3) the existence of valuable natural resources in the country; (4) the availability of cheap and skilled labor; (5) political stability, and so on. Incentives are generally mentioned way down the list. At the same time, attempts at removing the incentives often meet sharp reactions on the part of the investors and threats to move the activity to countries that offer incentives.

In a world of relatively homogeneous countries, the incentives provided by (only) one country would probably attract a lot of foreign investment to that country, especially in an open world in which companies can produce in one country and export what they produce to the rest of the world where import duties are lower than in the past. However, the world is far from homogeneous. There are great differences among countries in factors such as (1) inflation rates, (2) macroeconomic performance, (3) incidence of corruption, (4) stability and certainty of rule of law, (5) regulatory environments, (6) political stability, and (7) security. Often some of these factors overwhelm the advantages that incentives could provide to foreign investors in investing in specific countries. For this reason economists emphasize that the best incentives that countries can provide to investors are (1) a stable and good macroeconomic environment, (2) stable and clear regulations and rule of law, (3) tax rates that are competitive and tax systems that are not too complex, and (4) an environment that is relatively free of corruption.

An important element that often plays a major role in the willingness of countries' authorities to grant incentives for FDI is competition from other countries for FDI. This competition has been stim-

ulated by what is at times called "incentive shopping" on the part of foreign investors. Consider a country, say Costa Rica, that is approached by an American company, say Intel, and is told that the company is considering a large investment in Central America but that the specific location has not yet been chosen. Suppose the investment would create many good jobs and would contribute significantly to the country's economy. The foreign company demands substantial incentives from Costa Rica with the implicit threat that a refusal will cause the company to invest in a neighboring country so that Costa Rica would lose the investment. This is the situation that often confronts policymakers; they face offers that they find difficult to refuse.

This shopping for the best incentive package at times places smaller and poorer countries at a disadvantage and forces them to make concessions that can significantly reduce or even eliminate the benefits that a country can derive from foreign investment. A solution to this problem of incentive shopping by foreign investors could only come in the form of agreements, on the part of all the countries from a particular region, not to give in to such pressures and to establish jointly agreed on rules that restrict what the countries grant in terms of incentives to foreign investors. However, no such agreements exist and the desire to attract foreign direct investment continues to be a strong motivation for providing often too generous terms. The fact that some countries have developed at fast rates with the help of FDI remains a strong reason for policymakers to continue to provide generous incentives.

The foregoing analysis points to two basic questions. The first is whether subsidies and financial incentives can succeed in attracting FDI (or more FDI) to a specific country. The second is whether total FDI to countries can change due to incentives. Different strands of economic literature have dealt with these two questions.

The answer to the first question is probably yes, provided that the pull exercised by the existence of the incentives is not neutralized by other factors. Ample evidence from many countries suggests that they have succeeded in attracting large FDIs through the use of generous and well-targeted incentives (see Tanzi and Shome 1992). Other conditions being equal, incentives can attract FDI to a specific country. However, the situation implicit in such a ceteris paribus assumption rarely holds in the real world so that the value of incentives to foreign investors can easily be wiped out by some of the other factors mentioned earlier. A country such as Haiti will not become as successful as Ireland or Malaysia in attracting FDI strictly by offering incentives to foreign investors. Also, for every "winner" of a competition for FDI, there may be many "losers" that tried to attract FDI using subsidies (or other incentives) but were beaten by the better offer.

The answer to the second question is more difficult and on this issue the literature is more ambiguous. Already in 1962 Kaldor (1965) was raising the question and was giving an agnostic answer. There is no literature that has addressed specifically and directly this question. However, there is a lot of literature that has dealt with the closely related, even though technically different, question of whether taxes have affected the location of firms and investment (see for example Devereux and Griffith, 2002). This literature has generally concluded that "taxation matters," and it is easy to predict that it will matter more with the passing of time. However, the various studies have focused exclusively on taxation and have ignored the effects of other policies that may have biased the elasticity of investment decisions with respect to changes in tax rates. Furthermore it is still not clear how large an effect taxes have on FDI. It should be added that this literature normally discusses the location of FDI within OECD countries rather than the flow of FDI to developing countries that was the main focus of Kaldor's paper. Thus we are unable to state with any certainty whether developing countries, as a group, have received more FDI due to the incentives that they have been providing.

The Costs of Incentives What are the costs of the incentives to the countries that give them, especially to developing countries? Against the potential benefits in terms of employment, transfer of technology, improved productivity, and higher exports, there are costs, some obvious and some less so.

A first cost is the erosion of the tax bases. Incentives can sharply reduce the bases for the corporate income tax, the value-added tax, and the import taxes. At the same time the inflow of FDI may require additional public spending for infrastructures and other costs. If the FDI leads to a significant increase in the country's rate of growth, other taxes may go up and may compensate for the incentives given. But this is not always the case.

A second cost, and one much related to the first, is that the erosion of the tax bases caused by the incentives often forces the countries to raise the statutory tax rates on the sectors and enterprises that do not benefit from the incentives, thus reducing the growth potential of those sectors.

A third cost is that incentives inevitably create distortions in the allocation of resources. These distortions may be desirable if the incentives compensate the enterprises that receive them for positive externalities that they bring to the countries. This is for example the case when FDI brings new technologies that can be used by domestic firms. But this is not always the case, so the distortions created by favoring some firms over others can be economically costly.

A fourth cost is the one mentioned earlier connected with corruption. Incentives often become a fertile field for corrupt acts. Many studies have shown that corruption is costly to countries in the impact that it has on economic growth and also on equity.

Finally, there are potential costs associated with the competition in the granting of incentives. When countries compete for limited FDI and do so by progressively increasing the attractiveness of the incentives that they are willing to give, there is inevitably a race to the bottom that may be costly to all. There is also the possibility of a "winner's curse" where a country overestimates the value of an investment, pays too much, and winds up worse off despite succeeding in attracting the FDI.

In conclusion, in particular circumstances, well-thought-out and limited subsidies and financial incentives to FDI may be beneficial to countries that give them. However, more often than not, the results are not as good as hoped.

See also foreign direct investment (FDI); location theory

FURTHER READING

Bernardi, Luigi, Angela Fraschini, and Parthasarathi Shome. 2006. *Tax Systems and Tax Reforms in South and East Asia*. London: Routledge. Provides up to date de scriptions of the tax systems in Asian countries.

Billington, N. 1999. "The Location of Foreign Direct In vestment: An Empirical Analysis." *Applied Economics* 31 (1): 65 76. A useful analysis of factors that have influ enced the location of FDI.

Devereux, Michael P., and Rachel Griffith. 2002. "The Impact of Corporate Taxation on the Location of Ca pital: A Review." *Swedish Economic Policy Review* 9 (1): 79 102. A very good review of the impact of taxes on the location of investment.

Harrison, Ann E., and Brian Aitken. 1999. "Do Domestic Firms Benefit from Direct Foreign Investment: Evi dence from Venezuela." *American Economic Review* 89 (3): 605 18. Discusses potential externalities associated with FDI.

Kaldor, Nicholas. 1965. "The Role of Taxation in Eco nomic Development." In *Fiscal Policy for Economic Growth in Latin America* (Papers and Proceedings of a Conference held in Santiago, Chile, December 1962). Baltimore: The Johns Hopkins Press, 70 99. An early discussion by a major economist about the impact of taxes on economic decisions including capital flows.

Kenyon, Daphne A., and John Kincaid, eds. 1991. *Com petition among States and Local Governments: Efficiency and Equity in American Federalism*. Washington, DC: Urban Institute Press. A review of tax competition among American states.

Organisation for Economic Co operation and Develop ment (OECD). 2000. *Recent Trends in Foreign Direct Investment*. Paris: OECD. A useful reference with a lot of factual information on FDI.

Shah, Anwar, ed. 1995. *Fiscal Incentives for Investment and Innovation*. New York: Oxford University Press. Con tains several papers on the role of fiscal incentives.

Tanzi, Vito. 1995. *Taxation in an Integrating World*. Wash ington, DC: The Brookings Institution. An early book on the impact of globalization on tax systems.

Tanzi, Vito, and Parthasarathi Shome. 1992. "The Role of Taxation in the Development of East Asian Economies."

In *The Political Economy of Tax Reform,* edited by Ta katoshi Ito and Anne Krueger. Chicago: The University of Chicago Press, 31 61. A detailed review of the role of tax incentives in the development of Southeast Asian countries.

Tanzi, Vito, and Howell H. Zee. 2000. "Tax Policy for Emerging Markets: Developing Countries." *National Tax Journal* 53 (2): 299 322. A review article on tax policy and tax incentives in developing countries.

Zee, Howell H., Janet G. Stotsky, and Eduardo Ley. 2002. "Tax Incentives for Business Investment: A Primer for Policy Makers in Developing Countries." *World De velopment* 30 (9): 1497 1516. A comprehensive article on the role and the cost of tax incentives for investment.

VITO TANZI

■ Swan diagram

The Swan diagram is an analytical tool that facilitates streamlined analysis of macroeconomic policy for a small open economy. Developed in the 1950s by Trevor Swan (see Swan 1960; Salter 1959) for Australia, the Swan diagram was used to model how to achieve the two objectives of internal and external balance with two kinds of policy instruments, one affecting aggregate demand, the other a relative price. More recently, however, the Swan diagram has been generalized to other times, countries, and different sets of instrument variables. It facilitates policy analysis and policymaking by showing the kind of disequilibrium in the macroeconomy and therefore the required policy response. Shocks can also be analyzed according to the kind of disequilibrium they create.

A country attains *internal balance* when the aggregate demand for domestic output equals the aggregate supply of domestic output at full employment of resources and prices are stable. *External balance* requires equilibrium in the balance of payments. The current account surplus or the net balance of trade is exports minus imports. It must be financed by sustainable capital flows. If imports exceed exports, capital inflows are required; otherwise capital can be exported. Adjustment normally requires a change in relative prices combined with changes in demand or expenditure.

A key conceptual distinction for a small country is that it must take international prices as given. If markets are competitive, exportable and importable traded goods can be combined into a single category because a perfectly elastic world demand for exports and a perfectly elastic world supply of imports make their prices independent of domestic variables. But this means that the terms of trade, or the ratio of export to import prices, cannot change to help the adjustment to full equilibrium or balance.

Analytics Because the terms of trade are independent of domestic actions, the key relative price becomes the real exchange rate, Z, defined as the ratio of traded to *nontraded goods* prices (E P^*_T/P_N). Traded goods prices, P^*_T, are determined in world markets. Nontraded goods do not enter world trade; internal costs and demand determine their prices (P_N). Z is, therefore, the ratio of external prices to domestic costs. The nominal exchange rate (E) is measured in units of domestic currency per unit of foreign currency so that a rise is a devaluation or depreciation of the currency.

Domestic aggregate demand, or *absorption*, is total real expenditure on home produced goods and on imports and is the sum of personal and government consumption and investment (A = C + I + G). It must equal domestic output plus exports minus imports or the current account surplus. Therefore output minus absorption gives the current account surplus.

Consider a simple framework where Z and A are the two variables influenced by two policy instruments exchange rate policy, which changes E, and fiscal policy, which changes government expenditure, G. Fiscal policy directly affects absorption through government spending. If perfect capital mobility makes the domestic interest rate equal the international, and the nominal exchange rate is fixed, monetary policy has no freedom to change interest rates and affect domestic demand. If the nominal exchange rate, E, is fixed and world prices, P^*_T, are exogenous, the real exchange rate can change only with the nontradable goods price. E can, of course, be changed

through a periodic devaluation; if the exchange rate is managed more continuous changes are possible, and monetary policy can also affect aggregate demand.

The Swan diagram (figure 1) has the real exchange rate on the vertical axis and real absorption on the horizontal axis. The schedule IB_f gives internal balance or the combinations of Z and A at which output demand equals full employment output. The schedule is downward sloping because a rise in demand raises output and so does a rise in foreign prices, relative to domestic prices. As domestic absorption rises, Z must appreciate to reduce foreign demand for exports, and therefore total demand, to the full employment output level. Values above the schedule generate inflation as a more depreciated exchange rate and higher absorption raise demand; those below the schedule generate unemployment/ recession.

Schedule EB_0 gives external balance or the combinations of the two variables that yield an acceptable current account deficit. The EB schedule is upward sloping because the current account deficit worsens with A as imports rise, but a rise in Z improves it as exports rise and imports fall. As A rises Z must

depreciate to raise exports sufficiently to keep the current account deficit unchanged. The balance of payments is in surplus above the schedule since exports are higher and imports lower; it is in deficit below. The current account deficit is lower than capital inflows can safely finance above, and higher below, the EB schedule.

The point of full internal and external balance is F, where the two schedules intersect. These schedules divide the space into four zones of differing types of disequilibria. A combination of expenditure-changing and expenditure-switching policies is required to reach full balance F. The first changes the level of total demand in the economy. The second changes the direction of demand, shifting it between domestic output and imports and shifting domestic resources between traded and nontraded sectors. The latter is a necessary part of the adjustment process for a small open economy, where the real exchange rate can change only if the relative price of traded to nontraded goods changes. Switching effects are responsible for the slopes of the two schedules; otherwise they would be vertical. The shifting of resources on the production side occurs over a longer run com-
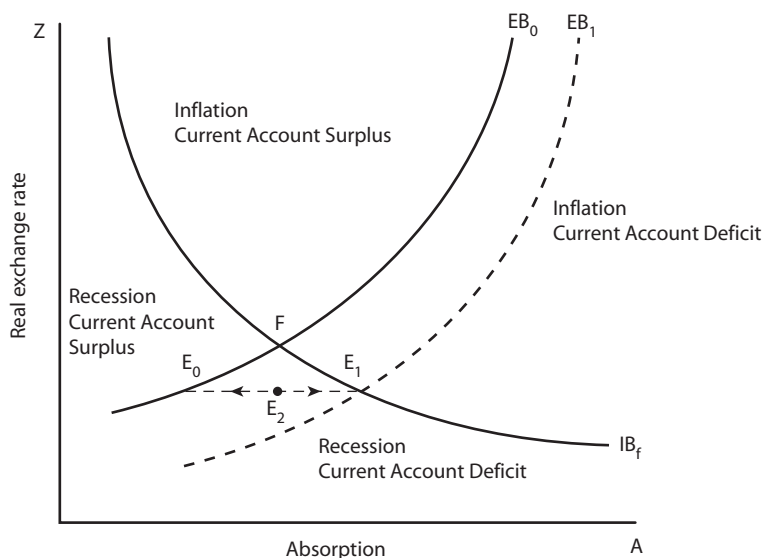
Figure 1
The Swan diagram

pared with the more immediate change in the direction of demand.

Domestic Costs and Wages If, as a simplification, labor is the only domestic resource, nontraded goods prices equal *unit labor costs,* or the nominal wage. Output supply depends on absorption and real wages and therefore on the real exchange rate. A rise in Z will make the production of export goods more profitable and labor will shift to this sector. In addition domestic demand will shift away from tradables, or the demand for imports will fall. But real wages themselves depend on the aggregate price level, which is a weighted average of the prices of traded and nontraded goods. Along EB, as Z rises, real wages must fall. If the price of the domestic good is closely linked to unit wage costs but wage earners spend on a weighted average of imported and nontraded goods, real wages must fall as imported goods become costlier than nontraded goods. Output is demand determined below the IB_f schedule and limited by available labor supply above it. Wage-price rigidities contribute to preventing labor market clearing.

Adjustment to Shocks Even if an economy is in full equilibrium a shock can place it in any one of the four quadrants where internal or external balance or both are missing. In analyzing adjustment to equilibrium it is useful to make a distinction between automatic forces of adjustment and policy responses. Prices of nontraded goods will rise with excess demand. This will reduce demand, but it will also appreciate Z, switch demand to tradables, and deteriorate the balance of payment. Over time labor will also shift to the production of nontradables, reducing the supply of exports and further deteriorating the balance of payments. Thus adjustment will be in the proper direction only in the top quadrant; in the quadrant to the right, the balance of payment will move further into deficit. Policy intervention, reducing government consumption of nontradables and depreciating the nominal exchange rate, would be required. A similar argument shows that deflation of prices and wages in the lower quadrants would be equilibrating for the balance of payment only in the bottom quadrant. In the quadrant to the left the surplus would tend to rise further, requiring policy

intervention that revalues the nominal exchange rate or raises government expenditure.

Depending on the slopes and speeds of adjustment, it can be optimal to assign one policy instrument to achieving internal balance and the other to achieving external balance. This is called the *assignment problem.* Moreover, two instruments are required to reach two objectives. For example, consider the economy to be at E_2, with the nominal exchange rate and real wages fixed so that the only policy instrument is government spending. If the government increases spending to reach internal balance at E_1, it moves further away from external balance. If it decreases spending, it can reach external balance at E_0 as imports fall, but recession deepens as it moves further away from internal balance. Even if there is no conflict and one instrument is pushing toward satisfying both objectives, changing the second may be necessary to reach the equilibrium point.

When deflation is required for automatic market-based adjustment, policy intervention is all the more necessary because deflation can be a very slow and painful process. In most economies it is easier to raise prices and wages than to reduce them. A rise in nominal exchange rates achieves the required adjustment in relative prices with less pain. But this only works if there is nominal but not real wage rigidity. Real wages have to be flexible for real exchange rates or the terms of trade to change. Nominal devaluation can be an effective policy instrument only if there is some flexibility in real wages. Nominal rigidity and real flexibility is the normal assumption in Keynesian models.

Swan originally analyzed what he called the "dependent economy." In the modern literature this has transmuted to "the small open economy," which also takes world prices as given. In this form it lives on with the *microfoundations* of forward-looking consumers who maximize welfare over their lifetime (Gali and Monacelli 2005). Wages and prices that take time to adjust to changed circumstances play a vital role in these models; these delays allow demand to have persistent effects. Prices continue to be closely linked to unit labor costs through markups, validating Swan's other fundamental contribution of the

importance of labor markets and domestic costs. The basic Swan diagram continues to be a useful simple approximation for policy analysis.

See also assignment problem; balance of payments; balance sheet approach/effects; currency crisis; exchange rate regimes; expenditure changing and expenditure switching; financial crisis; impossible trinity; Mundell-Fleming model; nontraded goods; real exchange rate

FURTHER READING

Corbo, Vittorio, and Stanley Fischer. 1995. "Structural Adjustment, Stabilisation, and Policy Reform: Domes tic and International Finance." In *Handbook of Devel opment Economics*, vol. 3, edited by Jere Behreman and T. N. Srinivasan. Amsterdam: North Holland, chapter 44. An application of the Swan framework to stabiliza tion and structural adjustment programs in the 1980s.

Corden, W. Max. 2002. *Too Sensational: On the Choice of Exchange Rate Regimes*. Cambridge MA: MIT Press. A readable and thorough application of the Swan frame work to derive policy lessons from the East Asian crisis.

Dornbusch, Rudiger. 1980. "Home Goods and Traded Goods: The Dependent Economy Model." In *Open Economy Macroeconomics*. New York: Basic Books, chapter 6.

Frankel, Jeffrey. 2001. "Coping with Crises in Emerging Markets: Adjustment versus Financing." In *Key Issues in Reform of the International Monetary System*, edited by P. Kenen and A. Swoboda. International Monetary Fund: Washington, DC, 215 20. Works out interest rate and exchange rate options during a crisis with the Swan framework in the presence of balance sheet effects.

Gali, Jordi, and Tommaso Monacelli. 2005. "Monetary Policy and Exchange Rate Volatility in a Small Open Economy." *Review of Economic Studies* 72 (3): 707 34. A rigorous New Keynesian model and policy analysis for the small open economy.

Goyal, Ashima. 2005. "Asian Reserves and the Dollar: Is Gradual Adjustment Possible?" *Global Economy Journal* 5 (3): Article 3, Downloadable from http://www .bepress.com/gej/vol5/iss3/3. Implications of labor markets and wage behavior for adjustment.

Krugman, Paul. 1998. "Latin America's Swan Song." Mi meo. Downloadable from http://web.mit.edu/krugman/ www/swansong.html. Working out Brazil's options with the Swan framework.

Rajan, Ramkishen S. 2006. "Managing New Style Currency Crises: The Swan Diagram Revisited." *Journal of Inter national Development* 18: 1 24. Builds on Frankel (2001) and works out interest rate and exchange rate options during a crisis with the Swan framework in the presence of balance sheet effects.

Salter, Walter E. 1959. "Internal and External Balance: The Role of Price and Expenditure Effects." *Economic Record* 35 (August): 226 38.

Swan, Trevor W. 1960. "Economic Control in a Dependent Economy." *Economic Record* 36 (March): 51 66.

ASHIMA GOYAL

target zones

See band, basket, and crawl (BBC); exchange rate regimes

tariff-cutting formulas

A tariff-cutting formula is an approach to trade liberalization in which countries bargaining over mutual tariff reductions seek agreement on a general formula for cuts in one another's tariffs, rather than on particular cuts in individual tariffs. Tariff-cutting formulas have been an inherent part of multilateral trade negotiations.

Economists are generally agreed that tariffs and other trade barriers reduce global income, and reducing these barriers will raise world income and the incomes of most countries. For small countries, economic theory provides an even stronger result, that unilateral liberalization will raise national income in all but the most exceptional cases. Despite these important and well-known conclusions, trade barriers remain extremely prevalent. Reconciliation of this apparent paradox needs to take into account the problem of special interests. Within a country, there is scope for these interest groups to collude with policymakers in order to promote their joint interests at the expense of overall welfare   perhaps by the interest groups making campaign contributions to politicians (Grossman and Helpman 1994). However, this outcome is suboptimal, even for politicians, for two reasons. One is the fact that the same decisions taken in partner countries reduce the demand for exports and shrink the supply of imports, causing the country's terms of trade to deteriorate (Bagwell and Staiger 1999). Another is that policymakers alone are unable to commit to rejecting future appeals for protection, allowing interest groups to apply intense pressure for protection   perhaps by threatening plant closures   that create costs rather than benefits for politicians (Maggi and Rodríguez-Clare 2007).

International trade negotiations provide potential solutions to these problems. Negotiations with trading partners can potentially deal with these problems by reducing foreign barriers in line with domestic barriers and providing a mechanism for policymakers to commit to not raising protection in the future. The GATT (General Agreement on Tariffs and Trade) achieves the second of these outcomes by international agreements limiting tariffs on individual products to no more than an agreed, bound tariff rate. Regional trade arrangements typically involve similar international commitments.

Such negotiations typically need to involve a significant share of a country's total trade if there is to be sufficient balance of gains to tackle the more politically sensitive trade barriers. One approach is for countries simply to negotiate bilaterally and sequentially, but this proved unsatisfactory during the 19th century, since a seemingly generous offer to one partner can be undercut by a slightly better offer to the next negotiating partner. The GATT attempted to improve on this through a request-and-offer procedure, under which the best offer given to any negotiating partner was extended to all GATT members under the so-called most-favored-nation

principle. By the 1960s, however, it was clear that this approach was yielding very little progress in terms of liberalization. The problem was that few bilateral trading relationships are intense enough that the gains in improved market access from an offer will offset the political costs involved in extending the offer to all GATT members when the number of participants becomes as large as the 74 participants involved in the GATT negotiations of the 1960s.

One solution, first adopted in the Kennedy Round (1963 67), was to negotiate a formula for cutting tariffs. In this situation, policymakers could potentially evaluate both the political costs of reducing their own tariffs and the political benefits from reducing tariffs in all of their trading partners. By agreeing on a proportional-cut formula during the Kennedy Round, participants in the GATT negotiations were able to achieve substantial cuts in protection 35 percent of initial tariffs as against an average of 2.5 percent in each of the previous four negotiations (Francois and Martin 2003). The next round, the Tokyo Round (1974 79), used a more sophisticated formula, the so-called Swiss formula, and achieved a 30 percent reduction in average tariffs. The Uruguay Round (1986 94) set broad tariff-reduction goals such as a 30 percent average-cut in agricultural tariffs, leaving the distribution of the cut across goods up to negotiations between trading partners, with a minimum cut of 15 percent for agricultural products. The Doha negotiations (2001), the first GATT round conducted under the World Trade Organization (WTO), have been focused on what are known as "Swiss formulas" for nonagricultural tariffs and tiered formulas, under which tariffs are grouped into four bands, with higher cuts on higher tariffs, for agricultural products.

The two classic tariff-cutting formulas the proportional cut and the Swiss formula are both extremely simple, and the contrast between them shown in figure 1 is illuminating. Under the proportional cut, the new tariff, t_1, equals $(1 - \alpha)$ times the original rate, t_0, where α is the fraction by which tariffs are to be cut. The Swiss formula is:

$$t_1 = \frac{a \cdot t_0}{a + t_0} \tag{1}$$

where a is a coefficient, or ceiling parameter, that no tariff will exceed after application of the formula. Figure 1 contrasts a proportional cut of 50 percent with a Swiss formula using a coefficient of 20 percent. Note that the proportional tariff-cutting rule results in a linear relationship between the original and final tariffs, while the Swiss formula is nonlinear. Note also that, for tariffs below the Swiss-formula coefficient of 20 in this case, the Swiss formula involves smaller cuts than the proportional cut of 50 percent. For tariffs above 20 percent, however, the Swiss formula involves larger cuts. These cuts become larger and larger as the initial tariff rises, in order that no postcut tariff will ever exceed the Swiss formula coefficient. This is evident from examination of equation (1), since $t_0/(a + t_0)$ comes closer and closer to 1 as the initial tariff rises toward infinity.

Progressivity A key difference between the approaches that have been used is the extent to which they are progressive in the sense of making the largest cuts in the highest tariffs. As is obvious from figure 1, the Swiss formula is sharply progressive. The proportional-cut approach cuts all tariffs by the same proportion, although this means larger absolute cuts in higher tariffs. The tiered-formula approach is somewhere in between, with higher cuts in higher tariffs. It is unlikely, however, that a tiered formula would be used to achieve such large cuts in very high tariffs as are implicit in the Swiss formula.
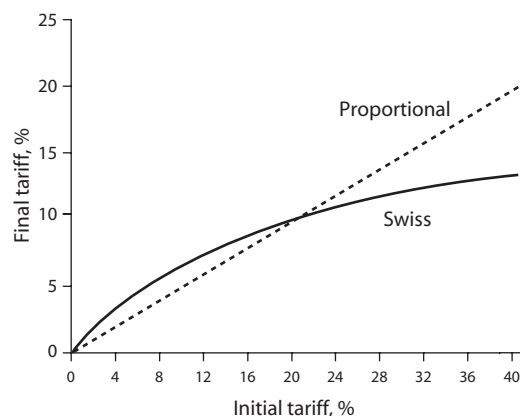
Figure 1

Proportional versus Swiss tariff cutting formulas

The average-cut approach, by contrast, tends to focus the largest cuts on the lowest tariffs. It rewards equally a 50 percent cut in a 1 percent tariff and a 50 percent cut in a 1,000 percent tariff   potentially allowing a 50 percent average cut in this case to be achieved by a 100 percent cut in the 1 percent tariff and certainly encouraging policymakers to focus their largest cuts on the lowest tariffs. This approach can result in a seemingly impressive tariff cut, with an agreed average cut of, say, 50 percent, that yields completely insignificant tariff reductions when economies have large numbers of low tariffs. A simple move from focusing on the average cut in tariffs to the cut in the average tariff could avoid this presumably unintended outcome.

Although there are few general rules about which approaches to tariff cutting will bring about the greatest improvements in economic efficiency, there is a general presumption that lowering higher tariffs by the most will be desirable. This follows from the simple rule that the economic cost of protection rises with the square of the tariff, implying that cuts in high tariffs are worth generally much more than comparable percentage cuts in low tariffs. The best approach depends on the goal of the negotiation, however. As Anderson and Neary (2007) have pointed out, when expanding market access is the goal, it may be more useful to focus tariff reductions on goods that already have low tariffs, since these goods tend to have the largest initial trade volumes. They show that, for a given mean tariff, reducing the variability of tariffs around their mean raises economic welfare but actually reduces the market access gains resulting from application of a formula.

The actual choice between different formulas appears, in practice, to depend a great deal on the structure of countries' tariffs. In the Kennedy Round, for instance, the United States resisted application of a progressive tariff formula since it had more high tariffs than the European Community. In the Doha negotiations, the Swiss formula appears to be the most broadly accepted option for nonagricultural tariffs. From the viewpoint of developing countries, this is partly because the industrial countries have only a small number of high tariffs, but many of these

are on products such as textiles, clothing, and footwear that are of great interest to developing country exporters. Since developing countries themselves have tariffs that are much less widely dispersed than those of the industrial countries, they are likely to have to make smaller cuts in their own tariffs than would be the case with a proportional cut that brought about the same reduction in global tariffs. From the viewpoint of the industrial countries, the ceiling feature of the Swiss formula would put limits on the extent to which tariffs could rise in the future in either developed or developing countries.

Flexibility A key difference between the different approaches used is the extent to which they impose disciplines on each tariff line. Line-by-line formulas such as the proportional cut or Swiss formula specify exactly what must be done with each tariff line. More flexible approaches, such as the average cut or an agreed cut in average tariffs, allow countries to choose what they do with individual tariffs subject to some overall constraint. Typically, when a line-by-line approach such as the Swiss formula is used, some flexibility is provided to allow for more lenient treatment for particular tariff lines. In the Kennedy and Tokyo Rounds, this was done by seeking as close as possible to full coverage, but ultimately allowing a substantial number of tariffs to be exempted from the formula treatment. In the Doha negotiations, constraints were being negotiated on the number of tariffs to be exempted from the full formula cut and the extent to which the cuts on these tariffs are allowed to deviate from the formula.

The ideal with such flexibilities is to provide the modest degree of flexibility needed to reach an otherwise unattainable agreement   a small hole in disciplines akin to that provided by the pressure valve on a boiler. The problem is that it is exceptionally difficult to know what represents such a modest degree of flexibility. What looks like a small hole, allowing dangerous pressures to escape, may turn out to be more like the small hole made in a balloon with a pin. Jean, Laborde, and Martin (2006) investigated the effect of allowing just between 1 and 5 percent of agricultural tariffs to be subjected to small tariff cuts instead of the full formula approach. They found that

allowing smaller cuts in just 2 percent of tariff lines was enough to reduce the cut in the average tariff resulting from the formula by more than two-thirds. Providing a balance between allowing the flexibility needed to get an agreement and the disciplines needed to make it worthwhile seems to require discipline on both the number of tariff lines and the degree to which the products subject to flexibilities are cut.

Generalization and Simplicity Trade negotiations tend to be very complex, even if they begin with very simple formulas such as the proportional cut formula. The proportional cut and Swiss formulas are extremely simple, involving only one parameter value to choose in each case. Both of these can be slightly generalized. The proportional cut can, for instance, be converted into a linear formula by adding an intercept term, b (see Panagariya 2002):

$$t_1 = b + (1 - \alpha) \cdot t_0 \tag{2}$$

This allows the marginal rate at which tariffs are cut to be higher than the average rate, perhaps to make the formula more progressive. One potential problem with this formula is that some initially low tariffs might actually be increased by a straightforward application of the formula.

The Swiss formula can be slightly generalized to allow for slightly smaller reductions in the highest tariffs balanced by slightly larger reductions in lower tariffs. Francois and Martin (2003) show that adding an additional flexibility parameter allows the shape of the formula to change in ways that could be useful in securing agreement. The resulting generalized Swiss formula is:

$$t_1 = \frac{a \cdot t_0}{a \cdot b + t_0} \tag{3}$$

A much more common approach to generalization has been to seek to identify different parameters that can accommodate the needs of different negotiating parties. The tiered formula under consideration in the Doha negotiations on agriculture has done this by including four tariff bands each for developed and for developing countries, with different cuts in tariffs within each band. This has resulted in an unfortunate level of complexity, with a total of 14 different parameters to be chosen—a level of specificity that leaves many unsure of the impacts of different choices, particularly the impact on their export market opportunities.

One area in which even the simplest nonproportional formulas, such as the Swiss formula, are much more complex in practice than a proportional cut is in the presence of tariffs that are expressed in non-ad-valorem terms. Such tariffs, which include specific elements such as a tariff of $1 per kilo, are very common in industrial-country agricultural tariffs. These present no particular problem when using a proportional cut formula. Fifty percent of a tariff of $1 per kilo is clearly 50 cents per kilo. But all the nonproportional formulas require that tariffs be converted into ad valorem terms before they can be applied. Such conversions are, in practice, potentially quite complex and subject to manipulation since it is not always clear what import price should be used to make the conversion. In the Doha negotiations on agriculture, months of intense negotiations were required before a compromise was reached on the approach to be used to make this conversion.

Tariff formulas can provide a very useful approach for negotiating reciprocal tariff reductions, particularly when the number of participants in the negotiations is large. They can potentially allow negotiators to secure agreements that would not be attainable by bargaining over individual tariff reductions. Popular, simple formula approaches such as the proportional cut and the Swiss formula were central to the success of the GATT in reducing industrial-country protection between the 1960s and the 1990s. The formula approach is not a panacea, however. In practice, the apparent need for flexibility to deal with particularly sensitive industries and products can result in complex and difficult negotiations about flexibilities and exceptions, which require careful management if negotiations are not either to collapse or to yield agreements that achieve very little.

See also General Agreement on Tariffs and Trade (GATT); multilateral trade negotiations; tariffs; World Trade Organization

FURTHER READING

Anderson, James E., and J. Peter Neary. 2007. "Welfare versus Market Access: The Implications of Tariff Structure for Tariff Reform." *Journal of International Economics* 71 (1): 187 205. Shows and measures the trade offs between tariff cutting formulas that increase market access and those that improve economic efficiency.

Bagwell, Kyle, and Robert Staiger. 1999. "An Economic Theory of GATT." *American Economic Review* 89 (1): 215 48. Provides a model to explain the success of the GATT as a means of reducing the costs imposed on trading partners by terms of trade deterioration result ing from tariffs.

Baldwin, Robert E. 1986. "Toward More Efficient Pro cedures for Multilateral Tariff Negotiations." *Aussen wirtschaft* 41 (2 3): 379 94. Shows the potential advantages of formula approaches and the limitations on the request and offer approach to securing trade liberalization.

Francois, Joseph, and Will Martin. 2003. "Formula Ap proaches for Market Access Negotiations." *World Economy* 26 (1): 1 28. Surveys some of the key formula approaches that have been used in tariff negotiations and proposes a generalization of the Swiss formula approach.

Grossman, Gene, and Elhanan Helpman. 1994. "Protec tion for Sale." *American Economic Review* 84 (4): 833 50. This classic article seeks to explain the structure of protection when politicians choose trade policies taking into account both voter well being and their ability to attract campaign contributions.

Jean, Sébastien, David Laborde, and Will Martin. 2006. "Consequences of Alternative Formulas for Agricultural Tariff Cuts." In *Agricultural Trade Reform and the Doha Development Agenda*, edited by Kym Anderson and Will Martin. Washington, DC: Palgrave Macmillan and the World Bank, 81 115. Examines the implications of tariff formulas under the Doha agricultural negotiations, taking into account the gaps between bound and applied tariffs and the presence of exceptions from the formula.

Maggi, Giovanni, and Andrés Rodríguez Clare. 2007. "A Political Economy Theory of Trade Agreements." *American Economic Review* 97 (4): 1374 1406. In tegrates the terms of trade explanation for international trade negotiations with the special interest model to provide a better understanding of the potential for trade reform through international trade negotiations.

Panagariya, Arvind. 2002. "Formula Approaches to Re ciprocal Tariff Liberalization." In *Development, Trade and the WTO*, edited by Bernard Hoekman, Aaditya Mattoo, and Philip English. Washington DC: World Bank, 535 39. An accessible introduction to the use of a wide range of potential formula approaches for tariff negotiations.

WILL MARTIN

■ tariff escalation

Tariff escalation refers to a situation in which tariffs are low or nonexistent for raw materials but increase, or escalate, with the degree of processing. If the tariff on cocoa beans is zero, for example, the tariff on cocoa butter is 10 percent and the tariff on chocolate and chocolate products is 20 percent. In the late 1950s and early 1960s, policymakers in developed countries considered escalating tariff structures modern and rational. Some economists, however, found that escalation in importing countries could discourage exports of processed products and thereby limit the scope for processing in exporting countries (see Johnson 1965). Developing countries became increasingly concerned with this problem and made the reduction or elimination of escalation one of their main demands in multilateral trade negotiations.

Tariff escalation has been part of the tariff struc tures of advanced industrial countries, as well as of those of many developing countries, at least since the 1960s. The reduction and elimination of high tariffs, tariff peaks, and escalation figure prominently in multilateral negotiations such as the World Trade Organization's Doha Round (still ongoing as of late 2007).

Measuring Tariff Escalation Tariff escalation can be measured as the difference among the nominal tariffs of products at different stages of the processing chain. Tariff escalation exists if the nominal tariff on the output exceeds that on the input; tariff deesca lation is present when the nominal tariff on the input

exceeds that on the output. The practical difficulty when using this approach is to describe processing chains and/or to identify products at different levels of processing.

Two main methodologies are typically used. The first uses a categorization of all products (tariff items) in a given nomenclature according to their degree of processing: typically raw materials, semifinished products, and finished products. The second requires the identification of processing chains for specific commodities, such as cocoa, cotton, or tobacco (OECD 1999). These two simple methodologies have limitations. They do not measure the intensity of protection afforded the final product, and they cannot be used when the final product incorporates a variety of inputs.

The concept of effective rate of protection (ERP) was devised in the early days of the theory of tariff structure to measure the effects of tariff structures on the allocation of resources. The ERP is defined as the percentage increase in value added per unit in an economic activity attributable to the tariff structure relative to the free-trade situation. It quantifies the intensity of protection afforded to the final product and can be used where multiple inputs are involved.

Consider the following simple example of a producer of bicycle frames that requires only steel as an intermediate input. Suppose that the production of a frame worth 150 in free trade requires the use of steel worth 100 in free trade. Value added at free trade prices is thus 50. Now assume that the country where the producer operates provides nominal protection at a rate of 20 percent on frames and 10 percent on steel. Under this structure of protection the domestic price of frames becomes 180 while the cost of the steel required to produce the frame becomes 110. Value added in this case is 70, which is 40 percent higher than value added in free trade. This increase in value added permitted by the protection structure is known as the *effective rate of protection*.

As shown in this simple example, in the presence of tariff escalation, the ERP (40 percent) is larger than the nominal tariff on the output (20 percent). Its main advantage is to capture certain general equilibrium mechanisms while avoiding the infor-

mational requirements of full general equilibrium computations. In the early 1970s, the utility of the concept of effective protection was challenged by a number of authors. Eventually, economists recognized the descriptive value of the ERP concept in empirical studies of protection as well as its analytical value subject to some caveats.

Effects of Tariff Escalation In principle, tariff escalation would be expected to distort resource allocation in both the importing and the exporting countries. In the importing country, output should be skewed toward processed products while in the exporting country it should be skewed toward raw materials. Yeats (1984), however, argues that the tendency for tariffs to increase over stages of a processing chain is neither a necessary nor a sufficient condition to establish a bias against processed goods. There are two problems: first, the nature of tariff-induced trade biases has never been defined precisely; second, the changes in demand conditions over stages of a processing chain have never been considered. Yeats shows that because import demand elasticities consistently increase with the degree of processing, neutral tariffs (i.e., tariffs that do not change the shares of primary and processed goods in exports relative to free trade while maintaining the aggregate level of imports) would often be deescalating. In other words, a trade bias against processed goods may exist even when actual tariffs deescalate over a processing chain.

In the 1990s, concerns were raised that tariff escalation may be a source of environmental damage to exporting countries. The argument was that by shifting economic activity toward primary production and away from processing, escalation induced excessive extraction of natural resources with consequent degradation of the resource base. A literature review by Hecht (1997), however, suggested that tariff escalation was not a significant source of distortion in the economic system or of damage to the environment. First, tariff escalation decreased with the Uruguay Round of international trade negotiations (1986 94) and was further reduced by preferences. Second, there was evidence that processing activities would tend to cause more environmental

harm in developing countries than in developed countries. Finally, the elimination of escalation would not boost income sufficiently in developing countries to significantly promote environmental protection.

Magnitude of Current Tariff Escalation In the last decade, a number of studies have assessed the incidence of tariff escalation in agricultural markets. These studies differ significantly in terms of methodology, country, and commodity coverage, as well as in the type of tariffs they consider. The studies generally point out the persistence of tariff escalation in the post Uruguay Round tariff landscape, and they provide a number of more specific results.

The Uruguay Round reduced escalation of developed country bound agricultural tariff rates. Even after full implementation of Uruguay Round tariff cuts, however, escalation persisted in developed countries in about half the agricultural product chains considered (Lindland 1997; UNCTAD 2003). USDA (2001) extends these results to developing countries. OECD (1999), which covers both agricultural and nonagricultural products as well as developed and developing countries, found that most processing chains in both members of the Organisation for Economic Co-operation and Development (OECD) and nonmembers are affected by escalation.

Applied tariffs also exhibit escalation. Elamin and Khaira (2004) show that for major developed importers, escalation prevails in 12 out of 16 of the agricultural product chains they consider. In developed countries, tariff escalation seems to be lower in the case of applied most-favored-nation tariffs than in the case of bound tariffs, while the opposite prevails in developing countries. When the nonreciprocal preferences that developed countries extend to developing countries are taken into account, escalation is clearly lower. In fact, the least-developed countries do not face any tariff escalation in markets where they are granted duty and quota-free access. It may be worth underlining, however, that nontariff barriers may also be escalating.

Since its effects on exporting countries were identified in the 1960s, escalation has been a concern for developing countries. Successive rounds of tariff negotiations have not gotten rid of it. How much the Doha Round of international trade negotiations will reduce escalation will largely depend on the degree of flexibility available to governments to protect specific products from formula cuts. In any case, for a reduction of escalation to have a substantial effect on the export structure of those countries that have not been able to diversify away from natural resources, most of which are least-developed countries, a number of other constraints on improved export performance will need to be released. Least-developed countries' poor export performance may be due in no small part to binding export supply response capacity constraints, which prevent fuller utilization of available market access opportunities.

See also commodity chains; effective protection; tariffs

FURTHER READING

Elamin, N., and H. Khaira. 2004. "Tariff Escalation in Agricultural Commodity Markets." *FAO Commodity Market Review 2003 4*, Rome: Food and Agriculture Organization of the United Nations, 101 20. A recent assessment of escalation in both bound and applied tariffs.

Hecht, J. E. 1997. "Impacts of Tariff Escalation on the Environment: Literature Review and Synthesis." *World Development* 25 (10): 1701 16. A review of the literature on the environmental effects of escalation.

Johnson, H. G. 1965. "The Theory of Tariff Structure with Special Reference to World Trade and Development." *Trade and Development*. Geneva: Institut des Hautes Etudes Internationales. One of the earliest analyses of the effect of escalation on exporting countries.

Lindland, J. 1997. "The Impact of the Uruguay Round on Tariff Escalation in Agricultural Products." FAO Com modities and Trade Division, ESCP No. 3. Rome: FAO. An assessment of tariff escalation in agricultural markets.

Organisation for Economic Co operation and Develop ment (OECD). 1999. *Post Uruguay Round Tariff Re gimes, Achievements and Outlook*. Paris: OECD. Includes an assessment of tariff escalation following the Uruguay Round in selected countries.

UN Conference on Trade and Development (UNCTAD). 2003. "Back to Basics: Market Access Issues in the Doha Agenda." UNCTAD / DITC / TAB / Misc. 9. Geneva: UNCTAD. An assessment of tariff escalation.

U.S. Department of Agriculture (USDA). 2001. "Profiles of Tariffs in Global Agricultural Markets." AER 796. Washington, DC: USDA. An assessment of post Uruguay Round escalation that covers developing countries.

Yeats, A. J. 1984. "On the Analysis of Tariff Escalation." *Journal of Development Economics* 15: 77 88. An inter esting discussion of the link between escalation and the existence of a bias against processed goods.

MARC BACCHETTA

■ tariff rate quotas

A tariff rate quota (TRQ) is unlike an ordinary quota because it does not set an absolute maximum level of imports. Instead, it sets a specified quantity of imports that can enter at a lower tariff (called an "in-quota tariff") than imports in excess of that limit. If the higher "over-quota tariff" is set so high that no imports occur above the limit, then the TRQ functions like an import quota. Historically the purpose of TRQs is to harmonize the common external tariff in a customs union or free trade area, implement tariff preferences for developing coun-tries, or provide a special safeguard against import surges (Rom 1979). More than 1,425 TRQs were established in the Uruguay Round Agreement on Agriculture to facilitate the conversion of import bans and quotas into tariffs (WTO 2006). In some cases, these agricultural TRQs increased trade be-cause countries were to provide "minimum access" (at least 5 percent of domestic consumption by 2001) and not restrict "current access" (quotas at historically established import levels). TRQs for agriculture are very important in members of the Organisation for Economic Co-operation and De-velopment, protecting upward of 50 percent of the total value of domestic production (de Gorter and Kliauga 2006).

Article XIII of the General Agreement on Tariffs and Trade (GATT) 1994 governs the administration of quantitative restrictions including TRQs and can be interpreted as inherently contradictory because it advocates nondiscrimination yet allows "supplier tariff quotas" as opposed to global quotas open to all exporting countries (Skully 2001). The GATT ad-vocates two criteria for proper administration: quota fill (where imports are to be determined by market conditions) and distribution of trade (trade patterns are to approach that without the restrictions). But importing countries have a GATT-consistent means of discrimination, because if agreement among all interested parties is "not reasonably practical," then quotas can be apportioned to exporting countries on a historical basis.

Disputes have resulted over the method by which country-specific export quotas are allocated. A classic example is the "Banana Dispute" (WTO 1997) where the method used by the European Union (EU) was deemed discriminatory because it did not reflect recent trade patterns. The WTO Appellate Body also ruled that the EU practiced discrimination in not requiring export licenses for all countries. The Ba-nana Dispute also highlighted problems in allocating import licenses because the EU was found to be in-consistent across countries in terms of the period of validity of the license, the size of license, eligibility requirements for a license, reallocation of unused licenses, and the requirements for the use of the li-cense. Some firms from Latin American countries faced complicated licensing procedures and had unnecessary burdens imposed on them. They were deemed to have been treated in a discriminatory and trade-distorting manner, so that the EU had to re-form its banana quota licensing system.

Ideally quotas should be auctioned, but mostly other forms of administration along with additional regulations have been imposed. The issue of TRQ administration is one of two major issues; the other is how best to liberalize trade with TRQs.

Trade Liberalization Effects of Reducing Tar-iffs versus Expanding Quotas Three basic "re-gimes" can determine imports and domestic prices, corresponding to each of the three policy instruments comprising a TRQ: the in-quota tariff (t_1), the over-quota tariff (t_2), and the quota level itself. The in-quota tariff regime can result for two reasons: market conditions are such that underfill occurs (the quota is not binding) or the quota (and over-quota tariff) is

disregarded, in which case either quota underfill or overfill can result. "Overfill" occurs when the government allows imports at the in-quota tariff to exceed the level of the quota. Imports paying the over-quota tariff (described later) are called "over-quota" and are not counted toward overfill.

Figure 1 shows the case of a small-country importer facing an import supply ES and an import demand curve ED such that the quota is not binding and underfill occurs. This shows that a quota fill rate of less than 100 percent does not necessarily imply inefficiency. But it could also be the case that the government allows the market to determine the equilibrium regardless of the quota level and so quota overfill could occur (not shown in figure 1), depending on market conditions and the levels of the quota and in-quota tariff. In either case, there are no quota rents. The domestic price Pd is equal to the world price inclusive of the in-quota tariff. Because the MFN over-quota tariff is the one under negotiation in the WTO, the difference between the over-quota and in-quota tariff represents "water" in the tariff, defined as that range over which a reduction in the over-quota tariff will have no impact on trade. An increase in the quota will have no impact initially with underfill due to market conditions and no impact at all if by government decree the quota is not a constraint.

The over-quota tariff determines the equilibrium in figure 2. One would normally expect the quota to be exactly filled, and imports beyond the quota to pay the over-quota tariff. But this particular case in figure

2 has quota underfill, which necessarily implies inefficiency in the administration of the quota licenses because some suppliers are forgoing quota rents in paying the higher tariff. Underfill in this case could be due to import quota licenses being allocated to higher-cost suppliers. Empirical research finds the situation depicted in figure 2 to be very common, but cases of exact fill and overfill also exist. Overfill can occur by government decree, when imports paying the lower tariff are extended beyond the quota while over-quota imports occur at the same time. In all cases, quota rents are positive, as are tariff revenues from both in-quota and over-quota imports. A reduction in over-quota tariffs will have an immediate trade liberalizing effect, while a quota expansion will have no initial impact because of over-quota imports and either quota underfill or overfill. Note that when a quota expansion does have an impact, there is a regime switch to the quota binding case (described next) but continued quota expansion will result in a switch to the in-quota tariff binding regime, in which case any further increases in the quota will have no trade liberalizing effects.

The case where the quota is binding initially is given in figure 3. The particular case shown has quota underfill, necessarily due to inefficient quota administration, but the cases of exact fill or quota overfill can occur as well. In all cases, there are positive quota rents and in-quota tariff revenues. A reduction in over-quota tariffs will have no immediate impact on trade because of water in the tariff due to
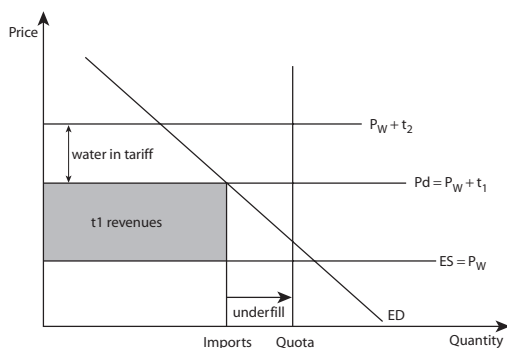
Figure 1

In quota tariff with quota underfill
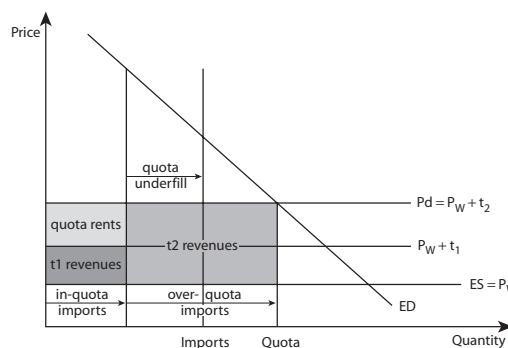
Figure 2

Over quota tariff with over quota imports and quota underfill
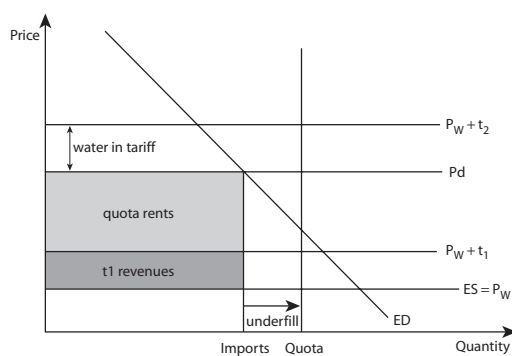
Figure 3
Quota binding with underfill

the binding quota. Expanding the quota will have an immediate impact only if there is exact quota fill initially; otherwise there will be no initial trade expansion with overfill and perhaps a less than one-to-one increase in trade with underfill, depending on the reason for the underfill.

The discussion shows that the efficacy of a tariff cut versus a quota expansion depends on a complex set of factors. The impact of tariff reductions depends on the level of water in the tariff, whereas that of quota expansion depends on the levels of over-quota imports, overfill or underfill, whether the in-quota tariff is initially binding, or how soon the in-quota tariff regime becomes binding with continued quota expansion.

TRQ Administration How the rights to quotas are distributed can affect the volume and distribution of trade, as well as the distribution of rents. With country-specific allocations, the exporting country may obtain the quota rents, depending on how the import licenses are awarded and the bargaining power between exporting and importing firms. The WTO (2006) identifies six principal methods of TRQ administration: applied tariffs; licenses on demand; first come, first served; historical imports; state trading enterprises (including producer groups); and auctions.

Applied tariffs refer to cases when the in-quota tariff determines imports and are usually used for less politically sensitive products. No licenses are allocated to importers   imports are not limited. There is

no inefficiency as a result except that due to the in-quota tariff itself.

With the *licenses-on-demand* method, firms request licenses and if the total licenses requested exceed the quota, then requests are reduced pro rata. Inefficiency is incurred because high-cost firms are allowed to operate closer to their optimal levels while low-cost firms are further from their desired levels. The ability to overbid exacerbates this inefficiency (Hranaiova, de Gorter, and Falk 2006).

Licenses allocated on the basis of *historical imports* also generate inefficiency if cost structures or demand change over time. Furthermore, firms may rent seek by importing at a loss at the over-quota tariff to increase their share of the licenses in the future.

The *first-come, first-served* basis hurries up imports and so causes inefficient allocation of imports over time. Because many agricultural products are perishable, rents are not dissipated with waiting in line. Instead domestic prices fall to world levels as exporters compete for the rent, which falls and so is appropriated by consumers. Domestic producers face the world price until the quota is filled, when the domestic price then spikes. The ability for firms to rent seek by storing the commodity or changing the seasonality of production will generate inefficiencies and reduce rent appropriation by consumers.

State trading enterprises (STEs) have direct control over imports; inefficiencies can result, depending on the STE's objectives and degree of control over imports and the domestic market. Some STEs in Asia tender licenses to the lowest-price bidder, resulting in low-quality imports (sometimes for animal feed) and so minimize the damage to producers for a given import quantity (Agricultural and Resource Economics Review 2000).

Because tariff quota administration is a rationing problem, economists favor a market-based system *auctioning* licenses to the highest bidder. Auctions account for the lowest level of agricultural trade under TRQs, however (de Gorter and Kliauga 2006).

Many additional regulations on import licenses can accompany each of these administration methods. The more onerous ones include domestic purchase requirements, time limits, limits per firm, and

seasonal licenses. A restrictive *domestic purchase requirement* can cause underfill, whereas *time limits* and *limits per firm* can impose increased costs of importation. *Seasonal licenses* increase inefficiency, especially in agriculture, where different harvest seasons cannot be taken fully advantage of between countries in the northern and southern hemispheres.

The role of TRQs in the modern world economy is mostly limited to agriculture, where they are facilitating the transition from nontariff barriers to tariffs as the principal means of protection. As of 2007, several outstanding issues in the WTO trade negotiations remain unresolved, including how best to liberalize TRQs. Reducing over-quota tariffs is generally agreed to be the more trade-liberalizing approach, but many countries instead expand quotas. This would reduce the amount of preference erosion for developing countries who often receive the quota rents. But it also leaves open the possibility of continued inefficiency due to the many ways quotas are administered, often with restrictive additional regulations.

See also distortions to agricultural incentives; nontariff measures; tariffs; World Trade Organization

FURTHER READING

Agricultural and Resource Economics Review. 2000. "Special Issue: Issues in the Administration of Tariff Rate Import Quotas in the Agreement on Agriculture in the WTO." ARER 20 (1): 54 124. A series of theoretical and empirical papers analyzing TRQs, with specific examples of country experiences.

de Gorter, Harry, and Erika Kliauga. 2006. "Reducing Tariffs versus Expanding Tariff Rate Quotas." In *Agricultural Trade Reform and the Doha Agenda*, edited by Kym Anderson and Will Martin. Washington, DC: World Bank and Palgrave Macmillan, 117 60. A comprehensive empirical analysis of TRQ liberalization in agriculture and the importance of alternative administration methods and additional regulations.

Hranaiova, Jana, Harry de Gorter, and James Falk. 2006. "The Economics of Administering Import Quotas with Licenses on Demand in Agriculture." *American Journal of Agricultural Economics* 88 (2): 3318 50. The only paper in the literature analyzing licenses on demand, the

TRQ administration method that has affected by far the most trade.

Rom, Michael. 1979. *The Role of Tariff Quotas in Commercial Policy*. New York: Holmes and Meier. An often cited, early analysis of the practice and theory of TRQs, including its history and treatment under the GATT.

Skully, David W. 2001. *Economics of Tariff Rate Quota Administration*. U.S. Department of Agriculture, Economic Research Service, Technical Bulletin Number 1893 (April). Washington, DC: USDA. One of the earliest systematic analyses of the economics of alternative TRQ administration methods.

World Trade Organization (WTO). 1997. *European Communities Regime for the Importation, Sale, and Distribution of Bananas*. Report of the Appellate Body (September 9). WT/DS27/AB/R. Geneva: WTO, Office of the Secretary General. A comprehensive account of the problems associated with the allocation of country specific quotas and import licensing schemes for EU banana imports, providing insightful details for researchers in understanding the variety of practical problems arising from the implementation of TRQs.

. 2006. *Tariff Quota Administration Methods and Tariff Quota Fill*. Committee on Agriculture Background Paper TN/AG/S/22 (April 27). Geneva: WTO. A comprehensive compilation of quota fill rates by quota administration method and additional regulation for different countries and commodity groups.

HARRY DE GORTER

■ tariffs

Tariffs are taxes levied on imported or, less often, exported goods when they cross the border between one customs territory and another. The use of customs duties, a synonym for import tariffs, can be traced back to the Babylonian civilization (18th to 6th centuries BCE). In the 20th century, tariffs fluctuated considerably. Since the end of World War II, however, thanks to eight rounds of tariff negotiations under the General Agreement on Tariffs and Trade (GATT) and to a number of preferential tariff agreements, developed country tariffs have been on a downward trend from an average of 20 30 percent to

between 3 and 7 percent. Developing country tariffs have also been substantially reduced, but only since the 1980s. Despite a substantial increase in the number of tariff lines, international comparability and overall transparency in the tariff area increased significantly with the adoption by all countries of a common nomenclature, the so-called Harmonized System (see Hoda 2001).

Because they are relatively easy to collect and seem to be paid by foreigners, tariffs have long been an instrument of choice for governments to collect revenue. In the late 20th and early 21st centuries, however, this role lost its importance among developed and most developing countries as governments became more interested in the effect of tariffs on the economy. Tariffs raise the price of imported products on the domestic market and reduce imported quantities. They yield revenue for governments and afford protection to domestic producers of the goods that compete with imports. A tariff imposed on some imported products will have two effects on the importing country: it will introduce welfare-reducing distortions of both production and consumption, and if the country that imposes it is large enough to affect foreign export prices, it will improve its terms of trade.

Types of Tariffs Tariffs are generally ad valorem, which means that they are expressed as a percentage of the value of imports, but they can also be *specific*, expressed as a number of monetary units per quantity unit (e.g., $5 per Kg). Other less common tariffs include *compound* tariffs, which have two components, one ad valorem and the other specific ($5 per Kg + 3 percent); *mixed* tariffs, which take the form either of a specific rate or of an ad valorem rate, whichever is the highest ($5 per Kg or 3 percent, w. i. t. h.); or *technical* tariffs, the level of which depends on the product's content of one or more specific input(s), such as sugar or alcohol. Non ad valorem tariffs are most frequent in agriculture but they are also used by a small number of countries for certain nonagricultural products.

An important distinction that relates to tariff negotiations and trade agreements is between most-favored-nation and preferential tariff rates. A basic

principle of the GATT and the World Trade Organization (WTO) is that governments should not set different tariff rates for different trading partners. This nondiscrimination principle takes the form of the most-favored-nation (MFN) clause, which dictates that a country must immediately extend any advantage, favor, privilege, or immunity granted to any product originating from or destined to any other country to the like product originating or destined to all other member countries. MFN tariffs apply to all trading partners with the exception of those that benefit from preferences. Multilateral rules allow exceptions to the MFN principle for regional trade agreements (GATT Article XXIV) and for non-reciprocal preferences granted to developing countries (1979 Enabling Clause). There are thus two different types of preferential tariffs: the nonreciprocal preferences granted by developed countries to developing countries, and the reciprocal preferences granted by participants in regional trade agreements to other participants.

From a legal perspective, a distinction is sometimes made between tariffs that are stipulated by laws, called statutory tariffs, and tariffs that are stipulated by conventions. The first category comprises, among others, the general rate that remains unchanged unless the situation substantially changes, temporary rates, preferential rates applicable to developing countries, and so on. Two types of duty rates are stipulated by conventions: conventional rates established in the WTO and preferential rates negotiated in regional trade agreements.

A second distinction that relates more specifically to the GATT/WTO is between bound and applied tariffs. When governments negotiate tariff reductions in the GATT/WTO, their commitments take the form of tariff bindings. MFN bound tariff levels, which are listed in a country's tariff schedule, indicate the upper limit at which the government is committed to set its MFN applied tariff. For a given tariff line, the MFN bound tariff must thus be higher than or equal to the MFN applied tariff, which is higher than the preferential tariff.

The concept of effective tariff or effective rate of protection serves to measure the encouragement

provided to domestic production by tariffs on both inputs and outputs. Nominal tariffs do not measure this encouragement appropriately. If they are imposed on imports that compete with a country's outputs, tariffs provide a positive stimulus to domestic producers. If, however, tariffs are imposed on a country's imported inputs, domestic producers are worse off. To capture the net protective effect of a tariff structure, effective tariffs measure the stimulus to value added, in the production of a particular product provided by nominal tariffs. The effective rate of protection is defined as the difference expressed in percentage terms between a sector's value added at world market prices and its value added with tariffs imposed. If the nominal tariff on inputs and output is the same, the effective tariff is equal to the nominal tariff. If the nominal tariff on the output exceeds the nominal tariff on the input, the effective tariff is higher than the nominal tariff.

Economic Effects of Tariffs Starting from a situation of free trade, consider the effect of a specific tariff t introduced by country A on product X. In the absence of any tariff or other distortion, the domestic price of product X is equal to the world price. When tariff t is introduced, however, exporters renounce shipping product X to country A if the price in A does not exceed the world price by t. The price in A starts rising while, if A's market is large enough, world price starts decreasing until the wedge is equal to t. The rise in price in A encourages domestic supply and discourages domestic demand. The volume of trade decreases. If A's market is large enough, the price increase in A is less than the amount of the tariff t. The decrease in A's demand lowers the world price so that only part of the tariff is passed on to consumers in A. If A is small enough, however, the world price remains constant and the tariff is fully passed on to consumers.

Consumers in A lose from the tariff while producers gain and the government collects tariff revenue. If A is large enough and the world price decreases, foreign producers lose and foreign consumers gain. A simple cost-benefit analysis using the notions of producer and consumer surplus reveals that for a country without market power the best

policy is free trade. The cost of the tariff to consumers is only partly compensated by the gain to producers and the increase in government revenue. The tariff induces a net welfare loss (deadweight loss) that can be linked back to both a consumption distortion (consumers do not consume enough X) and a production distortion (producers produce too much X). For a country that is large enough for changes in its demand to affect world prices, however, free trade is not necessarily optimal. In the large market case, the loss resulting from the two distortions must be compared with a "terms-of-trade" gain resulting from the lower price paid by consumers for the imported product. Note that this gain corresponds to a loss in the exporting country, which adds up to a deadweight loss resulting from the distortions of both consumption and production.

Arguments in Favor of Tariffs The cost-benefit analysis of tariffs has shown that for small countries without influence on their terms of trade, free trade is optimal. For countries that are large enough to influence the price of their imports, however, a nonzero tariff can be preferable to free trade on efficiency grounds. As first demonstrated by Bickerdike (1906), an "optimum tariff" can be calculated that maximizes national welfare. This optimum tariff level is the one at which the marginal benefit from improved terms of trade is equal to the marginal cost from production and consumption distortions. The so-called terms-of-trade argument for a tariff raises a number of empirical and practical problems, however. First, the practical relevance of the terms-of-trade argument is a matter of controversy among international economists. Second, as explained in the previous section, the terms-of-trade gain comes at the expense of the exporting country. If a large country uses its market power to extract gains from its partners, it will most likely induce retaliation and trigger a trade war (Johnson 1954).

Another argument that may justify the use of tariffs is the presence of market failures. If, for instance, capital or labor markets do not function properly and slow down or prevent resource reallocation, if activities of potential interest for production and export diversification generate learning

by doing spillovers, or if unemployment or under-employment is significant, the basic theoretical case for free trade does not hold. In the presence of market failures, the consumer and producer surplus concepts do not measure the costs and benefits of tariffs adequately. Consider the case where the production of some good or service generates productivity gains for the sector or even the whole economy and firms that produce the good cannot appropriate these benefits and thus do not take it into account in their production decisions. In such a case, firms will not produce the good despite the fact that it would be socially optimal to produce it. The spillovers could then be used as a justification for a temporary tariff that would stimulate production and thereby generate social benefits.

The market failure argument for protection has been criticized on two grounds. First, tariffs are not the first best response to most domestic market failures. In the spillover case, a production subsidy would also stimulate production, but unlike the tariff it would not distort consumption and would therefore be preferable. Second, the market failure argument in favor of tariffs has been criticized for practical and political reasons. For some economists, government failures cause more problems than market failures. Market failures are difficult to identify and appropriate policy responses difficult to design. Moreover special interests are likely to have more influence where governments pursue a more flexible trade policy.

Recent History By the end of World War I, high tariff levels, quantitative restrictions, and prohibitions combined with exchange controls dominated the trade policy scene. Countries that had become creditors to other nations, such as the United States, Argentina, Australia, Canada, and India, increased or kept tariffs very high by international standards. In the early 1920s, while various nontariff barriers introduced during the war were progressively lifted, tariffs were raised in compensation. In the late 1920s, a sharp fall in agricultural prices jeopardized attempts to negotiate tariff reductions. Germany, France, and Italy raised their agricultural tariffs and the U.S. Congress passed the protectionist Smoot-Hawley Tariff Act, which even before its implementation provoked widespread retaliation, contributing to a general increase in tariffs. The new trade restrictions aggravated the economic crisis and openly discriminatory trade relations evolved. In 1931, after almost one hundred years of a very liberal trade policy with minimal protection, the United Kingdom, still the largest importer, retreated via the Ottawa Agreement into a preferential trading system.

World War II severely affected the world trading system. A low level of trade went together with major trade imbalances. Trade policies were characterized by relatively high tariffs, the use of preferences by some countries, the extensive use of nontariff trade barriers by the European countries, and widespread government control of international transactions in order to manage scarce foreign exchange reserves of U.S. dollars. In 1947 the average tariff rate was between 20 and 30 percent (WTO 2007). The average U.S. tariff in 1947 was still considered to be among the highest in the major industrial countries despite a shift away from the extreme protectionist trade policies of the early 1930s. The average applied tariff rate among European countries ranged somewhere between 10 and 20 percent. A low-tariff country group (comprising Denmark, Norway, Sweden, and the Benelux countries Belgium, the Netherlands, and Luxembourg) had tariffs somewhat below 10 percent while a high-tariff group (comprising France, Italy, Portugal, and the United Kingdom) had tariffs averaging close to 20 percent (Woytinski and Woytinski 1955). The situation in other regions of the world was relatively diverse, but available evidence suggests that independent countries had relatively high tariffs. Brazil, for instance, had shifted toward a strict form of industrial protectionism in the late 19th century to become one of the most highly protectionist countries in the world.

Since 1947, eight rounds of multilateral tariff negotiations under the GATT and the WTO contributed to a substantial reduction of industrial tariffs in developed countries. Tariffs on agricultural and textiles and clothing products, which did not follow the downward trend, stick out above the average. Until the 1980s, a majority of developing countries pursued import-substitution policies with relatively

high tariffs. Starting in the 1980s, however, many if not most of them switched to more outward-oriented development strategies and started reducing their tariffs unilaterally, often in the context of structural adjustment programs.

At the same time, tariffs were also reduced on a discriminatory basis. In various regions of the world regional trade agreements were signed that resulted in the reciprocal elimination of tariffs on trade between a limited number of trading partners. Most developed countries also introduced nonreciprocal preferential tariff schemes aimed at encouraging imports from developing countries.

Priorities Because they have been brought down in most countries, tariffs are sometimes seen as history. In reality, however, they are still alive, and the evolution of the Doha Round of multilateral trade negotiations (ongoing as of late 2007) reflects how difficult it is for governments to agree to further cuts and bindings. There are a number of reasons for this. Tariff reductions tend to follow the path of least resistance, so the remaining duties can be expected to offer more resistance to liberalization efforts. It is not clear whether discriminatory tariff reductions facilitate or impede MFN reductions. More than anything else, however, skepticism regarding the benefits from tariff reductions in the public and in a number of governments has been growing. This may mean that the case for cutting and binding tariffs needs to be reassessed or simply that it needs to be made more convincingly.

See also effective protection; multilateral trade negotiations; nondiscrimination; tariff escalation; terms of trade; World Trade Organization

FURTHER READING

Bickerdike, C. F. 1906. "The Theory of Incipient Taxes." *Economic Journal* 16: 529 35. The earliest exposition of the terms of trade argument for a tariff.

Corden, W. M. 1974. *Trade Policy and Economic Welfare*. Oxford: Clarendon Press. A discussion of economic arguments in favor of and against protection.

Hoda, A. 2001. *Tariff Negotiations and Renegotiations under the GATT and the WTO*. Cambridge: Cambridge University Press. A discussion of procedures and practices in the implementation of the provisions relating to multilateral tariff negotiations and renegotiations.

Johnson, H. G. 1954. "Optimum Tariffs and Retaliation." *Review of Economic Studies* 21: 142 53. A demonstration of the fact that a country may gain by imposing a tariff even if other countries retaliate and a seminal contribution to the theory of trade agreements.

World Trade Organization (WTO). 2007. *World Trade Report*. Geneva: WTO. Chapter 2, "60 Years of International Trade Cooperation," includes a discussion of the history of tariff policies since World War I.

World Trade Organization, International Trade Center, and United Nations Conference on Trade and Development. 2007. *World Tariff Profiles 2006*. The most comprehensive source of aggregate tariff statistics.

Woytinski, W. S., and E. S. Woytinski. 1955. *World Commerce and Governments: Trends and Outlook*. New York: Twentieth Century Fund. Provides estimates of tariffs in the early 1950s.

MARC BACCHETTA

■ tariffs and foreign direct investment
See trade costs and foreign direct investment

■ tax holidays (or waivers)
See subsidies and financial incentives to foreign direct investment

■ Taylor rule
See monetary policy rules

■ technical barriers to trade
Standards and technical regulations exist to ensure consumer safety, environmental protection, national security, interoperability of electronic products, and many other goals. There is a wide variety of standards and technical regulations. These include guidelines for product certification, performance mandates,

testing procedures, conformity assessment rules, and labeling requirements, for example. The simple distinction between standards and mandatory technical regulations is that the latter refers to rules imposed by government, while standards are voluntary guidelines. Standards development may be encouraged by nongovernmental organizations, or required of a supplier to a manufacturer of a final good, for example.

From a producer's perspective, standards and technical regulations can overcome problems of fragmented markets and result in economies of scale. It is potentially more costly to produce for a market with many different incompatible products. This can deter firms from exporting to multiple markets with one product. Moreover, standards and regulations that promote compatibility or interoperability between products allow firms to supply goods at lower prices. From a consumer's perspective, efficient standards and regulations can make it easier to compare products to meet individual preferences and help to ensure that goods are safe or operate efficiently. Standards and mandatory regulations may also, however, act to restrict competition and act as a technical barrier to trade (Maskus, Otsuki, and Wilson 2005). This is particularly true when standards are nontransparent, imposed on a discriminatory basis, or not harmonized to international norms.

The majority of mandatory product regulations are set by national governments. Technical regulations may vary across countries given different views regarding optimal levels of safety, risk, levels of environmental protection, or national security, among other factors. Technical regulations that deviate from principles of reliance on international science, private sector led development, or international norms can raise costs for producers seeking to enter export markets and act as a technical barrier to trade. Technical regulations may discriminate against foreign suppliers, both in their design and implementation, and may be used to provide strategic trade advantages for domestic firms over foreign competitors. Technical regulations can also force firms to duplicate product testing and certification procedures required to test conformity to

standards and therefore raise design and production costs. Regulations can also be crafted to exclude foreign entrants into a particular market in order to support domestic monopolies.

Firms in developing countries may be especially affected by these factors, given less efficient production facilities, or lack of access to testing and certification procedures or information on international standards. The overall findings from the World Bank Technical Barriers to Trade (TBT) survey conducted in 17 developing countries indicate that 70 percent of firms that export must comply with mandatory technical regulations. The majority of firms responding to the survey perceive technical regulations to be important for entering export markets (Otsuki and Wilson 2004).

There are general commitments under the World Trade Organization (WTO) Technical Barriers to Trade Agreement that address the use of technical regulations to block trade (WTO 2007). These provisions encourage use of international standards to meet regulatory goals, promote mutual recognition of tests and certification where possible among members, and codify the principles of nondiscrimination and national treatment in regard to technical regulations. The TBT Agreement also mandates that governments notify the WTO of any new draft technical regulation prior to final implementation.

There is an increasing number of empirical studies that examine standards and technical regulations and their impact on trade flows. Among others, Chen, Otsuki, and Wilson (2006) examine how meeting foreign standards affects firms' export performance, reflected in export propensity and market diversification. The study uses firm-level data from the World Bank Technical Barriers to Trade Survey. The results indicate that firms subject to mandatory testing procedures have an export share nearly nine percentage points lower than other firms. The existence of information inquiry difficulties in major export destinations is estimated to decrease exports by 18 percent of their total sales. Furthermore, long inspection times for products significantly reduce firms' export incentives.

The study also explores how standards and technical regulations vary across different types of firms and how market diversification might be affected. Results suggest that the export propensity of domestically owned firms is affected more by testing procedures relative to other firms and that the negative relationship between informational inquiry difficulty and export propensity is strongest for manufacturing firms. In terms of market diversification, the results imply that standards reduce the likelihood for exporters to enter multiple markets.

Questions of the utility of harmonization of national standards and technical regulations to international norms are also increasingly important in regard to the role of standards as technical barriers to trade. Czubala, Shepherd, and Wilson (2007) examine the impact of European Union (EU) standards on African textiles and clothing exports, a sector of particular development interest. Evidence suggests that EU standards not harmonized to de facto international standards reduce African exports of these products. EU standards that are harmonized to ISO standards are less trade restricting. Results suggest that efforts to promote African exports of manufactures may need to be complemented by measures to reduce the cost impact of product standards, including international harmonization. In addition, efforts to harmonize national standards with international norms, including through the World Trade Organization Technical Barriers to Trade Agreement, promise concrete benefits through trade expansion.

Technical barriers to trade represent one of the more important potential obstacles to the expansion of world trade. Nontariff barriers have fallen over the past decades, and traditional nontariff barriers such as quotas and other policy tools of protection have been reduced. Efforts to deregulate, basing technical regulations on international scientific risk profiles, and reliance on suppliers' declaration of conformity reinforced by postmarket surveillance systems, among other steps, can help reduce the scope for governments to protect domestic firms behind technical barriers. Strengthening the WTO disciplines in the TBT Agreement, including stronger provisions on

notifications of new regulations by members and further encouragement of harmonization of standards to international norms, can also help.

See also nontariff measures; World Trade Organization

FURTHER READING

Chen, Maggie Xiaoyang, Tsunehiro Otsuki, and John S. Wilson. 2006. "Do Standards Matter for Export Success?" World Bank Policy Research Working Paper No. 3809. Washington, DC: World Bank. Examines the impact of standards and technical regulations on the propensity of firms to export.

Czubala, Witold, Ben Shepherd, and John S. Wilson. 2007. "Help or Hindrance? The Impact of Harmonized Standards on African Exports." World Bank Policy Research Working Paper No. 4400 Washington, DC: World Bank. Analysis in this paper draws on the first use of the World Bank EU standards database. It finds that, for African exporters, EU standards harmonized to International Standards Organization (ISO) standards in textiles and clothing are less trade restrictive than standards that are not.

Maskus, Keith, Tsunehiro Otsuki, and John S. Wilson. 2005. "The Cost of Complying with Foreign Product Standards for Firms in Developing Countries: An Econometric Study." World Bank Policy Research Paper No. 3590. Washington, DC: World Bank. Finds empirical evidence, based on the World Bank Technical Barriers to Trade database, that compliance with foreign product standards does involve costs to producers.

Maskus, Keith, and John S. Wilson. 2001. "A Review of Past Attempts and the New Policy Context." In *Quantifying the Impact of Technical Barriers to Trade: Can It Be Done?* Edited by Keith Maskus and John S. Wilson. Ann Arbor: University of Michigan Press, 1 7. Provides analysis of efforts to quantify the impact of standards and technical barriers on trade.

Otsuki, Tsunehiro, and John S. Wilson. 2004. "Standards and Technical Regulations and Firm's Ability to Export: New Evidence from World Bank Standards and Trade Survey." Mimeo. Washington, DC: World Bank. Provides a summary review of findings from the World Bank Technical Barriers to Trade database.

World Trade Organization (WTO). 2007. Technical Barriers to Trade: Technical Explanation. http://www.wto

.org/English/tratop e/tbt e/tbt info e.htm. Accessed June 2007. Provides an overview of rights and obligations of WTO members under the TBT Agreement.

JOHN S. WILSON

■ technological progress in open economies

Long-run economic growth is driven by technological progress, which in general takes two forms: product innovations and process innovations. The former refers to introduction of new or better quality products and the latter refers to implementation of resource-saving technologies. Diffusion of new products/technologies via imitation and direct technology transfer also promotes economic growth. This is especially valid for developing countries that are further away from the world technology frontier. The question of whether openness stimulates technological progress is important since openness is to a large extent determined by policy choices. Openness of an economy is measured by the degree of barriers to international trade, investment, and labor flows as well as by the volumes of these indicators.

Theory Openness can affect technological progress through a variety of channels. First, open economies can more easily acquire the state-of-the-art technologies available abroad via imports, foreign direct investment (FDI), and immigration. Second, exposure to international trade promotes competition among domestic firms. On the one hand, this can accelerate the pace of technological progress by motivating firms to escape competition via innovation. On the other hand, more intense competition can hurt profits, reduce the returns to research and thereby discourage innovation. Third, entrepreneurs in open economies have access to larger markets and thus enjoy larger returns on successful innovation. This raises the profitability of innovation and boosts growth. Fourth, open economies can avoid duplicating research conducted abroad and thus allocate their innovation resources more effectively. Fifth, under certain initial conditions, lowering trade barriers may reduce the relative price of skilled labor

and render research and development — which is a skilled-labor-intensive activity — less costly and thus stimulate growth. Finally, economies open to labor flows can experience higher growth by attracting human capital from abroad, which is known as brain gain. The opposite, brain drain, holds when human capital leaves the country. It should be noted that all of these mechanisms exert a short-run influence on economic growth but not necessarily a permanent influence on growth rates.

Case Studies Economists first investigated the linkages between growth and openness using case study techniques. These studies were conducted in the 1970s under the auspices of the World Bank, the Organisation for Economic Co-operation and Development, and the National Bureau of Economic Research (NBER). The basic methodology involved comparing the outcomes of import substitution (IS) strategies, as practiced in Argentina, Chile, Ghana, Mexico, Turkey, and other countries, with those of export promotion (EP) strategies, as implemented by Korea, Taiwan, Hong Kong, and Singapore. The premise of the IS policy was that developing countries need to protect their infant manufacturing industries by quotas and import tariffs in order to foster capital and technology accumulation. Proponents also suggested that by pursuing IS strategies developing countries would be able to avoid the adverse terms-of-trade effects associated with exporting raw materials. The premise of the EP policy was to provide market discipline to domestic firms by encouraging them to compete with foreign firms while keeping tariffs on imports at low levels. The case studies documented that pursuing EP strategies was a much more effective policy tool in fostering growth compared with IS strategies. These studies also underscored the need to complement trade liberalization with other fiscal, monetary, and structural changes such as stable government debt, market-determined exchange rates, and an improved education system.

Cross-Country Openness Growth Regressions Since the advent of broad international data sets in the early 1990s, economists have used cross-country growth regressions to investigate the openness-

growth link. The basic statistical methodology involves running regressions to identify the impact of openness on economic growth while controlling for other factors that are known to affect growth such as initial income per capita, investment in physical and human capital, political stability indicators, and so on. These studies can be classified into two categories based on their choices of openness indicators.

Trade volumes regressions. The first group of these studies has used total trade share, import share, and export share in gross domestic product as openness measures. In general, these cross-country growth regressions have established a positive and statistically significant link between growth and trade volumes. They have been criticized on two accounts, however. First, statistically significant correlations do not necessarily imply a causality from trade to growth. Second, trade is an endogenously determined variable that responds to income per capita. To tackle these issues, economists have used geographic indicators that is, variables that affect trade but are not influenced by policy and income as instruments and have obtained predicted values for total trade flows. They then regressed income per capita on these predicted values in the presence of control variables. The instrumental variable (IV) estimates have shown that the empirical findings from the standard cross-country growth regressions continue to hold to a large extent. The robustness of these IV estimates, however, have also been challenged by economists who incorporated institutional quality and geographic indicators in the regressions.

Trade restrictions regressions. The second group of econometric studies has used tariffs and nontariff barriers as openness indicators. These studies offer mixed results on the relationship between openness and growth and are plagued with measurement problems and data availability issues. These problems led economists to construct trade restrictiveness indexes that combine tariff and nontariff barriers with other indicators that are known to adversely affect international trade and investment. The most popular among these is the Sachs and Warner (SW)

index, which combines information on tariff rates, quotas, political control of exports, exchange rate distortions, and market structure. Although many studies have found this index to have a positive correlation with economic growth, the index is too broad to interpret as a measure of trade or investment policy. Further, it was shown that when one uses specific trade-related subcomponents of the SW index in particular tariff rates and nontariff barriers the relationship between openness and economic growth becomes statistically insignificant.

As an alternative to cross-country studies, recent work has focused on within-country effects of trade liberalization. The methodology involved comparing the pre and post trade liberalization growth rates using econometric techniques. One study that determined the timing of trade liberalization based on the SW index calculates that on average countries that switch from being closed to being open experienced a large and statistically significant increase in their growth rates. Obviously, not all of the gain in growth can be attributed to opening up the economy, because in most cases trade liberalization is accompanied by other policy changes that may foster growth such as financial reform, privatization, and investment in infrastructure.

International Technology Diffusion via Imports and FDI Empirical studies have also examined the magnitude of technology diffusion via imports by accounting for the stock of knowledge embodied in imported products. These studies show that knowledge-intensive imported goods in particular differentiated capital goods and high-technology imports exert a large and significant influence on domestic productivity. This is consistent with endogenous growth theory, which predicts that openness increases access to new technologies and specialized high-quality intermediate inputs and thereby stimulates productivity growth.

Technology spillovers can also take place through FDI and multinational firm (MNF) activity. MNFs can bring new technologies that can be diffused to the local economy via labor training, labor turnover, and provision of high-quality intermediate products. Interactions between MNFs and domestic firms in

the form of vertical production relationships can further contribute to technology diffusion. A number of industry- and plant-level studies have found that the presence of foreign-owned firms leads to technology spillovers, but the nature and magnitude of these spillovers may differ across countries. For instance, a study focusing on Venezuela finds that the intensity of foreign ownership adversely affects the productivity of domestic firms. Another study focusing on the United States reports that 11 percent of U.S. manufacturing productivity growth is accounted for by inward FDI to the United States. The number of industry- and plant-level studies has been quite limited so far, but recently there has been a call for such studies and more work appears to be on the horizon. In the meanwhile, cross-country growth regressions have found that the growth-promoting effects of FDI materialize only for countries that attain a threshold level of human capital and financial development.

Implications for Policy Even though substantial empirical evidence supports the view that trade promotes growth, most economists caution that it is not necessarily and exclusively policy-induced trade that serves as the magic bullet. Changes in national trade and investment policies bring forth a reallocation of resources within an economy. Growth-promoting effects of trade liberalization can be reaped fully only when complemented with appropriate fiscal, monetary, and structural policies that facilitate the allocation of resources to their most productive use. These include but are not limited to stable government budget positions, low inflation, market-determined exchange rates, an improved education system, a well-functioning legal system, and reduced corruption.

See also capital accumulation in open economies; economic development; foreign direct investment and international technology transfer; growth in open economies, Schumpeterian models; intellectual property rights; international income convergence

FURTHER READING

Baldwin, Robert E. 2003. "Openness and Growth: What Is the Empirical Relationship?" NBER Working Paper No. 9578. Cambridge, MA: National Bureau of Economic Research. Provides an overview on the theory and empirics of openness growth relationship.

Frankel, Jeffrey A., and David Romer. 1999. "Does Trade Cause Growth?" *American Economic Review* 89 (3): 379 99. The first model to investigate the impact of trade on income per capita levels taking into account the endogeneity of trade.

Keller, Wolfgang. 2004. "International Technology Diffusion." *Journal of Economic Literature* 42 (3): 752 82. Provides an extensive survey of technology diffusion across borders through investment and trade.

Helpman, Elhanan. 2004. *The Mystery of Economic Growth*. Cambridge, MA: Belknap Press. Provides a summary of the evolution of growth theory with a chapter on the role of interdependence in affecting growth.

Rodriguez, Francisco, and Dani Rodrik. 2001. "Trade Policy and Economic Growth: A Skeptic's Guide to Cross National Evidence." In *NBER Macroeconomics Annual 2000*, edited by Ben Bernanke and Kenneth Rogoff. Cambridge, MA: MIT Press, 261 325. Provides a critical evaluation of the cross country studies on growth and trade.

Wacziarg, Romain, and Karen Welch. 2003. "Trade Liberalization and Growth: New Evidence." NBER Working Paper No. 10152. Cambridge, MA: National Bureau of Economic Research. Empirically investigates how trade liberalization changes the pattern of growth within a country.

M. FUAT SENER

▪ technology licensing

Technology licensing occurs when an economic agent with a particular technology sells that technology to another economic agent without it. Depending on the characteristics of the licensed technology, technology licensing may either reduce the licensees' cost of production or increase the quality of the licensees' products.

The literature on technology licensing can be separated into two categories based on whether the licenser and the licensees compete in the product market. Whether the licenser and the licensees are

from different countries may have also important implications.

Kamien (1992) surveys the work on patent licensing in cases where the licenser and the licensees are not competitors in the product market. This entry, in contrast, will concentrate mainly on cases where the licenser and the licensees compete in the product market.

Since licensing improves the licensees' technology, it affects the intensity of product market competition and the profits of the licenser and the licensees, thus changing the firms' product market strategies (such as output or pricing decisions), which may have important welfare implications.

Optimal Licensing Contract and Quality of the Technology Considering licensing between two final goods producers, Rockett (1990) derives optimal two-part tariff licensing contracts, which consist of an up-front fixed fee and a per-unit output royalty. Royalty-only licensing is the optimal contract if there is no threat of imitation from the licensee and the licensee's output is verifiable. If imitation is costless to the licensee, however, the optimal licensing contract involves only an up-front fixed fee. Combining a fixed fee and a royalty is optimal if the threat of imitation is credible and the cost of imitation is positive.

With regard to the quality of the licensed technology, where a lower marginal cost of production represents relatively better technology, the optimal decision of the licenser is not to license its best possible technology if the threat of imitation is credible and the cost of imitation is moderate.

Since output royalty softens competition in the product market, a licenser who competes in the product market with the licensee may find it optimal to charge a positive royalty rate even if the royalty rate distorts the licensee's output. If the products of the licenser and the licensee are horizontally differentiated (e.g., by different brand names or after-sales services of the firms), however, the incentive for fixed-fee licensing increases since product differentiation itself softens competition between the firms. Whether product differentiation increases the threat of imitation and the incentive for licensing a rela-

tively better technology is ambiguous. Product differentiation increases the incentive for imitation for a given royalty rate by increasing the gain from imitation, but product differentiation also reduces the royalty rate, thus reducing the incentive for imitation.

If the products are homogeneous, the optimal licensing contract may involve only an up-front fixed fee if the high cost of either technology licensing or technology adoption makes it possible to license the technology to only some competitors. Stealing market share from the nonlicensees makes fixed-fee licensing optimal.

The royalty rate does not have a competition-softening effect if the licenser and the licensee do not compete in the product market. As shown in Gallini and Wright (1990), however, there still may be a rationale for charging a positive royalty rate if the quality of the technology is proprietary information of the licenser, as the royalty rate helps to signal the quality of the technology. If there is licensing between competitors from different countries, the presence of trade costs provides a rationale for a combination of a fixed fee and a royalty.

In Rockett (1990), licensing intensifies competition in the product market and may induce the licenser not to license its best technology. Kabiraj and Marjit (1993) argue that even if the licenser and the licensee do not compete in the product market in the absence of licensing, a licenser may have an incentive not to license its best technology if the licenser and the licensee are from different countries. Government policy may prevent competition in the licensee's market; the licenser's having better technology than the licensee may prevent competition in the licenser's own market. If licensing the best technology would create competition in the licenser's market, the licenser may find it optimal not to license its best technology.

Licensing and Other Product Market Strategies While technology licensing allows a firm to increase its profit, other business strategies such as mergers, joint ventures, and capacity commitments also help firms to increase their profits.

Consider the impacts of product market competition on a bilateral licensing contract and a bilateral

merger. If there are multiple technologically superior and technologically inferior firms, a licensing contract always dominates a merger in a Cournot oligopoly. If there are multiple technologically inferior firms and only one technologically superior firm, however, merger dominates licensing for a sufficiently large technological difference. Under merger, the merged firms lower their total outputs compared with no merger, which, in a Cournot oligopoly, induces the nonmerged firms to increase their outputs. Hence merger creates a positive externality on the nonmerged firms and reduces the benefit of merger. On the other hand, licensing increases output of the licensee by reducing its marginal cost of production. Hence in a Cournot oligopoly, licensing reduces output of the licenser and also steals market shares from the nonlicensed firms. Therefore, along with the benefit of cost efficiency of the licensee, stealing of market shares from the nonlicensed firms provides the incentive for licensing. The trade-off between the positive externality of merger on the nonmerged firms and stealing of market shares from the nonlicensed firms is responsible for these results. Roy et al. (1999) extend this line of research in open economies and show the relative profitability of licensing, joint venture, and cross-border merger.

In the case of international technology transfer, proprietary information about the quality of the foreign technology and imitation by the licensee are two common problems. Though strategic choice of the fixed fee and output royalty may help to resolve these problems, a joint venture between the licenser and the licensee may induce relatively better technology to be transferred compared with licensing.

If imitation under international technology transfer creates the threat of competition in the licenser's other product markets, a technologically superior firm may choose foreign direct investment (FDI) over licensing. Though FDI requires significant investment in the host country, it may help to protect the licenser's other markets by reducing imitation by host-country firms. Hence the patent policy of the host country may influence a foreign firm's incentive for licensing versus FDI, with FDI more likely to be observed under weak patent protection.

Instead of a cooperative strategy such as merger, a noncooperative strategy such as investment in capacity or incentive delegation to managers may also help to raise the profits of the firms. Using licensing with an up-front fixed fee, Mukherjee (2001) shows that capacity commitment and incentive delegation have significantly different effects on licensing. Whether both the licenser and the licensee or one of them has the opportunity to adopt these precommitment strategies also plays an important role.

Licensing and Innovation Differences in technologies, which create reasons for licensing, stem from innovation. The interaction between innovation and licensing is an important topic.

In a duopoly market structure, where the firms innovate and compete in the product market, Gallini and Winter (1985) consider licensing ex ante and ex post innovation. The trade-off between the replacement of inefficient production technology and the elimination of inefficient research expenditure shows that licensing encourages innovation if the firms' initial technologies are close in costs, but it reduces innovation if the costs are asymmetric.

Katz and Shapiro (1985) consider separately an innovator and producers who can be engaged in licensing and show that licensing between the producers reduces innovation if the licensee appropriates most of the gains from licensing. If the licensee appropriates a significant amount of gains from licensing, the incentive for getting property rights to the innovation is reduced, thus reducing the innovator's earnings and the incentive to innovate.

These two papers have important implications for innovation policies, but they are unable to show the effects of licensing on R&D organization. It is often found, however, that firms do cooperative R&D to reduce the cost of innovation, avoid duplication of research efforts, or both. Whether licensing increases the incentive for cooperative R&D depends on the motive for cooperative R&D and the nature of the product market competition.

Even if licensing does not affect innovation, it may affect the decision on technology adoption,

which occurs after innovation. Given that a technology has a finite life span, Mukherjee and Pennings (2004) show that licensing (compared with no licensing) induces technology adoption earlier by eliminating imitation.

Glass and Saggi (2002) show the effects of licensing and FDI on innovation and growth in an open economy. There is a cost disadvantage under FDI, whereas licensing may require the firms to share the gains from licensing. If the mode of operation (i.e., licensing or FDI) of the foreign firm is fixed, a rise in the licensee's share of rents is detrimental for innovation and economic growth. A mode switch from licensing to FDI increases innovation and economic growth on average, however.

Welfare-Reducing Licensing The reason for encouraging technology licensing is its positive effect on welfare, since it replaces relatively inefficient technologies, thus eliminating production inefficiencies. Recent research shows, however, that one must be cautious in encouraging licensing. Since licensing makes the licenser and the licensee more symmetric, it helps to sustain collusion between the firms. This argument was first put forward by Eswaran (1994), who showed that cross-licensing of products between firms can be a collusive device.

Licensing of a superior process technology also helps to sustain tacit collusion between the licenser and the licensee by making them symmetric in terms of costs.

Faulí-Oller and Sandonis (2002) show that licensing may be welfare-reducing even in the absence of collusion. They show that if the firms compete on prices, licensing may reduce welfare by contracting output and thus raising product prices. If the firms compete on quantity and the licenser can commit its output before the licensee, the provision of output royalty in the licensing contract helps to reduce industry output and to raise the product price, and thus may reduce social welfare.

Another strand of literature shows that licensing may reduce welfare by affecting either an innovator's incentive to innovate or the firms' incentives for doing cooperative R&D.

In the presence of multiple licensees, if the technologically superior firm can choose a licensing contract contingent on the number of firms taking licenses, the licenser can earn the profit of a monopolist with the licenser's technology. Hence if the licenser can use a more complete contract in the sense that it does not need to follow a per-unit output royalty and can charge nonlinear royalties, it may reduce welfare by increasing market concentration.

Licensing to Create Competition A surprisingly common feature of most research on licensing is the assumption that input markets are perfectly competitive. Hence it is assumed that technology licensing, while affecting the licensees' marginal cost by making superior technology available to them, does not affect the licenser's marginal cost. It is often the case, however, that input markets are imperfectly competitive.

One such possibility arises when there are labor unions bargaining over wage rates. In this case, a monopolist final goods producer may license its technology to a potential competitor to create competition in the product market. While licensing increases competition in the product market, it may help to reduce the wage rate charged by the labor unions. Since a suitably designed royalty rate softens competition between the firms, the benefit of lowering the wage rate induces the monopolist final goods producer to create competition through licensing.

Even if there are no labor unions but a monopolist final goods producer is selling to a foreign country and is exposed to the host-country tariff policy, the monopolist may be better off creating competition through licensing. Whether the monopolist licenses the technology to another foreign firm (which also faces the tariff rate of the host country) or to a host-country firm may be a decision variable of the foreign monopolist.

Farrell and Gallini (1988) and Shepard (1987) show the rationale for licensing by a monopolist input supplier, which is a well-known phenomenon in the Ethernet market and the semiconductor industry. Licensing by the monopolist input supplier to a potential competitor creates competition in the

input market and helps to reduce the input price, which in turn, increases profitability of the final goods producers. Hence higher competition in the input market due to licensing encourages entry in the final goods market, thus increasing the demand for inputs, and makes the monopolist input supplier better off compared with no licensing. Even if licensing in the input market does not affect the entry decision in the final goods market, it may increase the demand for inputs by raising the (expected) input quality, thus making the monopolist input supplier better off under licensing than no licensing. Higher competition in the input market due to licensing helps to commit to a lower input price or to a higher input quality, thus increasing the profit of the monopolist input supplier.

Future Research Significant effort has been devoted to several issues related to technology licensing; at least two areas deserve more attention, however. First, the literature has mainly focused on licensing of cost-reducing innovations, but firms often innovate to improve the quality of the product and/or to invent new products. Second, the licensing literature ignores licensing by several technologically superior firms. The presence of multiple licensers will affect market outcomes such as the terms of the licensing contracts, the quality of the licensed technology, and social welfare. Work by Saggi and Vettas (2002) that considers the equilibrium contracts when two upstream firms contract with multiple downstream firms may be reinterpreted as technology licensing by multiple technology producers to multiple final goods producers. However, this area requires further research.

See also foreign direct investment (FDI); foreign direct investment and international technology transfer; foreign direct investment under oligopoly; foreign market entry; intangible assets; intellectual property rights and foreign direct investment; internalization theory; intrafirm trade; joint ventures; outsourcing/offshoring; proximity-concentration hypothesis; technology spillovers

FURTHER READING

Eswaran, M. 1994. "Cross Licensing of Competing Patents as a Facilitating Device." *Canadian Journal of Economics* 27 (3): 689 708. Shows that cross licensing of products can facilitate tacit collusion.

Farrell, J., and N. T. Gallini. 1988. "Second Sourcing as a Commitment: Monopoly Incentives to Attract Com petition." *Quarterly Journal of Economics* 103 (4): 673 94. Shows optimality of technology licensing by a mo nopolist input supplier due to licensing.

Faulí Oller, R., and J. Sandonis. 2002. "Welfare Reducing Licensing." *Games and Economic Behavior* 41 (2): 192 205. Examines welfare reducing licensing under price competition.

Gallini, N. T., and R. A. Winter. 1985. "Licensing in the Theory of Innovation." *RAND Journal of Economics* 16 (2): 237 52. Looks at the effects of licensing on the incentive for innovation.

Gallini, N. T., and B. D. Wright. 1990. "Technology Transfer under Asymmetric Information." *RAND Journal of Economics* 21 (1): 147 60. Argues that optimal licensing contract involves both fixed fee and output royalty in presence of private information about the quality of technology.

Glass, A., and K. Saggi. 2002. "Licensing versus Direct Investment: Implications for Economic Growth." *Journal of International Economics* 56 (1): 131 53. Shows the effects of licensing and foreign direct invest ment on innovation and economic growth.

Kabiraj, T., and S. Marjit. 1993. "International Technology Transfer under Potential Threat of Entry A Cournot Nash Framework." *Journal of Development Economics* 42 (1): 75 88. Explores why, in an open economy, a licenser may not find it optimal to license its best technology.

Kamien, M. I. 1992. "Patent Licensing." In *Handbook of Game Theory with Economic Applications*, edited by R. J. Aumann and S. Hart, 331 54. Amsterdam: Elsevier Science B. V. (North Holland). Surveys the works on licensing where the licenser and the licensees do not compete in the product market.

Katz, M. L., and C. Shapiro. 1985. "On the Licensing of Innovations." *RAND Journal of Economics* 16 (4): 504 20. Shows the effects of licensing on innovation when the innovator and the producers are different firms.

Mukherjee, A. 2001. "Technology Transfer with Com mitment." *Economic Theory* 17 (2): 345 69. Shows the

effects of capacity commitment and incentive delegation to managers on technology licensing.

Mukherjee, A., and E. Pennings. 2004. "Imitation, Patent Protection, and Welfare." *Oxford Economic Papers* 56 (4): 715 33. Examines the effects of licensing on tech nology adoption and welfare.

Rockett, K. 1990. "The Quality of Licensed Technology." *International Journal of Industrial Organization* 8 (4): 559 74. Looks at the effects of imitation on optimal licensing contract and the quality of the licensed tech nology.

Roy, P., T. Kabiraj, and A. Mukherjee. 1999. "Technology Transfer, Merger, and Joint Venture: A Comparative Analysis." *Journal of Economic Integration* 14 (3): 442 66. Shows the effects of cost asymmetry in an open economy.

Saggi, K., and N. Vettas. 2002. "On Intrabrand and In terbrand Competition: The Strategic Role of Fees and Royalties." *European Economic Review* 46 (1): 189 200. Characterizes equilibrium contracts with an up front fixed fee and output royalty when the upstream firms contract with multiple downstream firms.

Shepard, A. 1987. "Licensing to Enhance the Demand for a New Product." *RAND Journal of Economics* 18 (3): 360 68. Shows the optimality of licensing by a monopolist input supplier due to the effect of licensing on the quality of input.

ARIJIT MUKHERJEE

■ technology spillovers

Technology spillovers are the beneficial effects of new technological knowledge on the productivity and innovative ability of other firms and countries. Technology is "nonrival": one's use of a technology does not limit its use by others, and the cost for an additional agent to use an existing technology is negligible compared with the cost of inventing it. Hence not all the benefits of technological knowledge are appropriated by the inventor; technological investments typically generate social returns that far outweigh private returns. Technology, once invented, can be used and diffused internationally with small added cost but substantial added benefit.

Technological research and innovation is mostly undertaken by firms and governments in the leading world economies that are also the world technological leaders. Then technology diffuses to the rest of the world through the main channels of trade, migration, foreign direct investment (FDI), and technological licensing (patents and copyrights).

International technology spillovers have received much attention in recent economic research from both theoretical and empirical perspectives. Theory identifies them as a key mechanism for the sustained growth of productivity and its diffusion across countries. From an empirical point of view, economists have studied how to measure technology spillovers and what channels are conducive to them. From a policy point of view, countries desiring greater technology spillovers use policies to promote trade and FDI and to promote better conditions for taking advantage of spillovers by absorbing them into domestic productivity gains.

Theory of Technology Spillovers Recent theories of economic growth and income differences across countries (see Eaton and Kortum 2002; and Klenow and Rodriguez-Clare 2005 for reviews) identify available technological and scientific knowledge as the most important determinants of productivity in a country. Scientific and technological innovation are the main engines of productivity growth in the rich countries (Europe, Japan, and North America). Their diffusion to industrializing countries, accompanied by investments in physical and human capital, is the main reason for the growth in productivity and income per capita of those economies. Yet some countries seem stuck far behind the technology frontier. The process of technological diffusion has a central position in the recent literature on development and growth. A better understanding of the nature of technology spillovers should help shed light on why some countries grow faster than others.

Due to its nonrival nature, technological knowledge can be used by producers other than the inventor to increase their productivity. Hence it generates two types of benefits called "spillovers."

1. First, new technological knowledge can be used in any country to produce more

efficiently or higher quality goods. This spillover increases the labor productivity of the country that adopts it.

2. Second, technological knowledge can be used in any country to produce new ideas or new applications in research and development (R&D). This increases R&D effectiveness in receiving countries.

Inventors usually appropriate at least part of the benefits from the first type of spillovers, either by producing goods with the new technology and exporting them to foreign markets (trade) or by setting up production that uses the new technology in other countries (FDI) or by licensing out the new technology and receiving royalty payments for it. International trade, FDI, and international patents and copyrights are therefore common channels for diffusing the benefits of technological innovations to consumers in other countries. At the same time those flows carry the knowledge related to the new technology embodied in goods, in machines, or in instructions. This new knowledge enables receiving countries to benefit from the second type of technological spillovers: other firms and producers may learn and improve their productivity as a consequence of exposure to better technology. The first type of technology spillovers are usually mediated by market mechanisms (trade, investments, and intellectual property rights) and are sometimes called technology diffusion. The second type of technology spillovers involve diffusion of knowledge to other firms of the receiving country via mobility of workers, learning, imitation, and subcontracting and are considered technology externalities.

In the light of these beneficial effects on productivity and growth, international technological spillovers (even more than international trade of goods and international movements of capital per se) have been identified as potentially the most beneficial aspect of globalization. The empirical research has consequently focused on measuring the intensity, quantifying the effects, and identifying the most relevant channels of these spillovers. At the same time researchers and policymakers have analyzed what are the characteristics that make a receiving country best

positioned to receive the benefits from those spillovers.

Empirical Analysis of International Technology Spillovers Technology spillovers are not recorded in the data. The channels of their transmission (trade, FDI, patents) and their consequences (productivity benefits) can be recorded and measured, however. The recent empirical analysis has used a plurality of data and approaches to qualify and quantify the intensity and productive impact of those spillovers across countries.

Measuring Technology Spillovers Technological and scientific knowledge is an intangible asset not measurable directly. Economists have used measures of R&D resources (input) or measures of innovations such as patents or productivity (output) to approximate it. Aggregate studies have used country-level data, while microlevel studies have used firm-level data. Two general methods are used to identify spillovers. The first method considers the effects of R&D done in some countries (or firms) on the productivity of other countries (or firms) that are linked to the former via trade, FDI, or technological/geographical proximity. The basic features of this approach were first developed in a very influential paper by Coe and Helpman (1995) and will be described in the next section. The second approach considers directly the association between the presence/intensity of trade and FDI (channels of technology spillovers) and the productivity of the importing/receiving country or firms there. Both methods infer the existence of spillovers indirectly from the effects on productivity in firms of the receiving economy.

A third approach aimed at identifying more directly the linkages that reveal technology spillovers analyzes citations from a patent to previous patents, considering them as tangible signs of a knowledge spillover (Jaffe and Trajtenberg 2002). An existing idea recorded in a patent contributes to the development of a new idea (new patent), and the citation link reveals this spillover. This method isolates only spillovers of the second type described above (from R&D to R&D) and tends to emphasize the geographic localization of those technology

spillovers. This method complements (but cannot substitute for) the other type of studies, as it only identifies the intensity and characteristics of technological spillovers but cannot quantify their impact on productivity.

Trade and Technology Spillovers A popular approach for analyzing the presence and the intensity of technology spillovers has been to analyze the association (correlation) between productivity in country j (or industry or firm) and the R&D activity in countries (industries or firms) other than j that are linked to j by potential channels of spillovers. The basic empirical procedure, presented in Coe and Helpman (1995), and expanded and updated since then, is to estimate some version of the following regression: Productivity$_j$ = Function($\mathbf{X}_j$, R&D$_j$, R&D$_{spillovers}$). The variable Productivity$_j$ represents some measure of the productivity (usually total factor productivity or labor productivity) of country (industry, firm) j. The vector $\mathbf{X}_j$ is an array of country (industry, firm) characteristics relevant for its productivity. R&D$_j$ is a measure of research and development activity performed in the country (industry, firm) j, and R&D$_{spillovers}$ = $\sum_{i \neq j} m_i(\text{R\&D})_i$ is a weighted sum of R&D activity in other countries. This term captures technology spillovers because the weights m_i are constructed to reflect the intensity of potential spillover channels between country (industry, firm) j, the receiver, and country (sector, firm) i, the sender. In the original work of Coe and Helpman (1995) m_i was measured as the share of imports from country j among trade partner of country i assuming that imports are the most relevant channel of technology spillovers. Subsequent research has experimented with different weights to capture other potential spillover channels. Some alternative measures of m_i have been the share of FDI from country j in total capital formation of country i, the share of imports of capital goods (rather than all goods), and the share of direct and indirect trade (i.e., trade through third-country mediation). More recently Keller (2002) has constructed the weights m_i based on geographical proximity between country j and i. This approach is based on the assumption that a whole array of potential spillover channels

(trade, FDI, migration, technological licensing, business travel, and others) are strongly enhanced by geographical proximity. A popular variation of the approach described above, mostly used on individual firm data, is to measure the effect of technological spillovers on production costs rather than productivity.

The limit of this type of approach is that the identification of externalities is based on correlations, and it is not easy to establish a real causation link. To address the limits of the reduced form approach, recent research by Eaton and Kortum (2002) has studied the relationship between R&D technology diffusion and domestic productivity in the context of general equilibrium models. In those models one can analyze and simulate the impact of increased research and trade liberalization on technology spillovers and productivity.

Overall the findings of this literature point to two rather robustly estimated regularities. First, the effect of R&D spillovers on productivity is consistently larger than zero and significantly positive for the average analyzed country (usually the data cover economies that are members of the Organisation for Economic Co-operation and Development, OECD). Second, while for the leading economies (G7 countries), the impact of domestic R&D on productivity is consistently larger than the spillover effects from other countries, for smaller and less advanced economies (other OECD countries) the impact of spillovers is larger than the impact of domestic R&D on productivity. These findings confirm that technological leaders tend to perform most of the R&D and innovation in the world and that spillovers from those technologies are a major source of productivity growth in other countries.

FDI and Technology Spillovers A second approach to identifying and quantifying international technology spillovers is based on the idea that FDI is an explicit activity set up to transfer technology across national borders (e.g., Markusen 2002). Hence FDI is a direct carrier of technology flows. The question is how much these flows benefit the productivity of the receiving economy and what are the features or policies of the receiving economy that enhance the

positive effects of technology spillovers. Several theoretical models argue that multinational enterprises should generate technology spillovers to local firms through several channels, and many of them have been studied in detail using firm-level data. Imitation, learning, and acquisition of human capital through worker turnover are considered as the most important channels of spillovers. Competition, subcontracting, and supply of high-quality intermediate inputs are market-mediated mechanisms that make better inputs available to the local firms, stimulate more efficient technologies, and may also have positive productivity effects.

The typical empirical approach for identifying technological spillovers through FDI estimates the following model: Productivity$_{jk}$ = Function($\mathbf{X}_j$, FDI$_k$). The term (Productivity$_{jk}$) measures the productivity of firm j in sector k and the right-hand side of the expression implies that it is a function of a vector of firm characteristics $\mathbf{X}_j$ and of some measure (usually the share of employment or of sales) of the presence of multinational enterprises in sector k, (FDI$_k$). Usually these studies analyze firm-level data for one country at a time (hence no country subscript) and often they consider geographical proximity as a requisite for (or enhancer of) the technology spillovers: multinationals have larger productivity effects if they are located in the same region or area as the potential spillover-receiving firm j. Another relevant dimension of the spillovers is whether they benefit domestic firms "horizontally" (namely, those in the same industry) or "vertically" (those that supply inputs to or buy inputs from the multinational enterprise).

Using cross-section and panel data from several different industrialized and industrializing countries, many researchers have estimated technology spillovers through FDI using the method described above. Interestingly, the evidence in favor of positive effects of technology spillovers on productivity of domestic firms is scant, especially when considering developing countries. Blomström and Kokko (1998) and more recently Görg and Greenway (2004) review dozens of such studies and conclude robust evidence of positive effects of FDI spillovers is found

mainly for industrialized countries. For developing countries results are much less clear. A typical example of those studies is the influential article by Aitken and Harrison (1999), which, in analyzing evidence from FDI in Venezuela, does not find any positive effect on productivity of local firms. In fact, the study finds some negative effects on domestic firms and attributes them to increased competition from FDI and crowding out of domestic firms.

Absorptive Capacity These findings have prompted research on the role of the receiving firm or country in determining the impact of technology spillovers. While the potential for technology spillovers is intrinsic to FDI activity, the actual impact on productivity of domestic firms depends on the absorptive capacity of those firms. Human capital and investment in R&D by the receiving country (firm) are important prerequisites to generating positive effects of technology spillovers on productivity. Insufficient human capital in local firms could be the explanation for the lack of spillovers from FDI in developing countries. Using firm-level data, several articles (e.g., Glass and Saggi 1998) confirm that firms using low-skilled workers and backward technology are unable to benefit from FDI technology spillovers.

Policy Considerations FDI is seen by governments of several industrializing countries as highly beneficial to the domestic economy. Technology spillovers are only one of its effects, as FDI also increases employment opportunities, consumption, and government income. Nonetheless technological spillovers generate probably the most important and lasting effects in the long run as they channel technological transfer and induce productivity growth. In the light of the empirical findings, policies promoting higher education and skill-formation of workers and R&D investment of domestic firms are needed complements to policies that attract FDI, if a country is to maximize the absorption of technological spillovers. At the same time some policies such as R&D requirements, technology transfer requirements, and local hiring targets for multinational enterprises can increase the intensity of technological activity and the benefits for the local economy.

Probably, however, a general framework of openness to international flows and free competition in trade, FDI, and technological licensing is the most important component for a country to attract technology spillovers and benefit from improved technologies as they become available to the global economy.

See also agglomeration and foreign direct investment; appropriate technology and foreign direct investment; foreign direct investment (FDI); foreign direct investment and international technology transfer; foreign direct investment and tax revenues; foreign direct investment under monopolistic competition; foreign direct investment under oligopoly; foreign market entry; intangible assets; intellectual property rights and foreign direct investment; internalization theory; joint ventures; linkages, backward and forward; location theory; proximity-concentration hypothesis; technology licensing

FURTHER READING

Aitken, Brian, and Ann Harrison. 1999. "Do Domestic Firms Benefit from Foreign Direct Investment? Evidence from Venezuela." *American Economic Review* 89 (3): 605 18. One of the most cited and influential articles on FDI spillovers in developing countries.

Blomström, Magnus, and Ari Kokko. 1998. "Multinational Corporations and Spillovers." *Journal of Economic Surveys* 12 (3): 247 77. An early comprehensive survey of the literature on FDI and technology spillovers.

Coe, David, and Elhanan Helpman. 1995. "International R&D Spillovers." *European Economic Review* 39 (5): 859 87. The initial and still one of the most influential articles on trade and technology spillovers.

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography, and Trade." *Econometrica* 70 (5): 1741 79. A sophisticated theoretical and empirical analysis of technological diffusion and technology spillovers.

Glass, Amy, and Kamal Saggi. 1998. "International Technology Transfer and the Technology Gap." *Journal of Development Economics* 55 (2): 369 98. One of the most influential recent articles on the role of absorptive capacity in determining the impact of FDI spillovers.

Görg, Holger, and David Greenway. 2004. "Much Ado about Nothing? Do Domestic Firms Really Benefit from Foreign Direct Investment?" *World Bank Research Observer* 19 (2): 171 97. The most comprehensive recent review of the role of FDI in promoting technological spillovers.

Jaffe, Adam, and Manuel Trajtenberg. 2002. *Patents Citations and Innovations: A Window on the Knowledge Economy.* Cambridge, MA: MIT Press. The most comprehensive collection of research on the use of patent data and patent citations to measure technology spillovers.

Keller, Wolfgang. 2002. "Geographic Localization of International Technology Diffusion." *American Economic Review* 92 (1): 120 42. A very influential paper measuring the importance of geographical proximity for technological diffusion.

———. 2004. "International Technology Diffusion." *Journal of Economic Literature* 42 (3): 752 82. A comprehensive literature review on international technology diffusion.

Klenow, Peter, and Andres Rodriguez Clare. 2005. "Externalities and Growth." In *The Handbook of Economic Growth*, edited by Aghion Philippe and Steven Durlauf. Amsterdam: Elsevier. An excellent review of the most relevant theoretical models of technological spillovers and economic growth.

Markusen, James. 2002. *Multinational Firms and the Theory of International Trade*. Cambridge, MA: MIT Press. An influential, mostly theoretical book on multinational firms and international trade

GIOVANNI PERI

■ **temporary movement of natural persons**

The phrase *temporary movement of natural persons* (TMNP) refers to one of several ways by which, according to the General Agreement on Trade in Services (GATS), international trade in services takes place when workers from one country travel abroad temporarily to supply particular services there.

The exact World Trade Organization (WTO) definition of TMNP can be found in Article I 2 (d) of the GATS, which refers to "the supply of a service, . . . by a service supplier of one Member, through the presence of natural persons of a Member in the territory of any other Member." This

definition is further qualified in the GATS Annex on the Movement of Natural Persons Supplying Services under the Agreement, which states that GATS rules, negotiations, and commitments do not apply to persons seeking permanent citizenship, residence, or employment. TMNP in the strict WTO sense applies only to temporary migration of workers in service occupations and not to foreign workers employed temporarily in the manufacturing, mining, or agricultural sectors of another country. National immigration policy in most countries, however, covers temporary workers in different sectors of the economy. For example, U.S. visa rules on temporary foreign workers include an H2 visa category for temporary farm workers.

TMNP thus implies at least two things: first, the overseas travel by the worker has to be *temporary* in nature; second, it must have been undertaken in order to perform a *specific job or supply of a service*. Examples include senior managers posted abroad to manage operations of a multinational firm operating in the services sector of a foreign country; IT technicians traveling overseas to provide backup technical support for computer hardware or software; academic faculty traveling to foreign universities on short-term teaching assignments; nurses or teachers employed in foreign hospitals or educational institutions for a fixed tenure; seafarers working on contract in a foreign shipping line; household workers hired for a specific period as au pairs by families abroad; an accomplished doctor traveling to a foreign country to perform complicated surgery; cultural troupes performing overseas; and unskilled and semiskilled workers from one country employed on contract on construction sites in other countries.

Examples of TMNP that do not strictly fall within the GATS definition, but are perhaps as numerous and of equal economic significance include farm labor employed annually on agricultural lands and plantations overseas during peak seasons; foreign workers in major mining projects; and workers employed in labor-intensive manufacturing activities during, for example, an export boom in a foreign country. Several bilateral and some regional agreements cover such TMNP that takes place outside the services sector.

Temporary overseas travel that is *not* considered TMNP includes travel undertaken for tourism or for obtaining medical assistance or education. Such journeys, though recognized as contributing to international trade in services, are classified under a different category, namely "consumption abroad," in the GATS. The movement of refugees seeking temporary asylum in other countries is also excluded, as travel is not undertaken for a specified period or to perform a specific task.

Origins Temporary migration has been under discussion for a long time in various forums such as the International Organization for Migration and the International Labor Organization but only as a subset of migration in general. The WTO was the first international organization to focus exclusively on the temporary migration of workers. Furthermore, while other organizations use the conventional word *workers*, the WTO introduced the phrase *natural persons* instead, in order to distinguish people from juridical persons such as companies or organizations. This merely reflects a preference for legalistic terminology by the drafters of the GATS text, however; in practice, the terms are interchangeable.

The origins of the concept can be traced to the Uruguay Round trade negotiations (1986 94) of the General Agreement on Tariffs and Trade. During discussions to develop a framework for an agreement on international trade in services, it was recognized that one way in which services are delivered overseas is through the travel of workers. Although overseas travel to supply particular services was not a new phenomenon traders, religious missionaries, and diplomats had been doing so for centuries there had been a sharp rise in such travel during the preceding decades. Lower transportation costs, widening income differentials between countries, and global distribution of production chains facilitated by technological improvements in communications were some factors responsible for this trend. Aging workforces and income growth have also led to worker shortages in some countries, creating a de-

mand for foreign workers. The emergence of significant numbers of temporary migrant workers worldwide by the 1980s appears to have contributed to the inclusion of TMNP as a negotiable trading activity during the Uruguay Round.

Leading TMNP Countries Reliable data on TMNP flows are not readily available. This is mainly because many countries do not keep records of persons going abroad for work. In host countries the statistics are complicated by the fact that many workers admitted temporarily get themselves converted into permanent workers.

Traditionally Mexico and the Philippines have been regarded as the world's leading source countries for TMNP. Because of the previously mentioned data limitations, some experts prefer to use workers' remittance figures as an indirect measure of TMNP, given that temporary workers tend to remit a significant part of their earnings to their home countries relative to permanent migrants. Measured in terms of the magnitude of remittances received in 2005, the top 10 countries sending temporary workers abroad were India, China, Mexico, the Philippines, France, Spain, Belgium, United Kingdom, Germany, and Egypt, in that order.

A similar indirect measure can be used for estimating the leading host countries for temporary migrant workers. This makes use of the data on wages and salaries earned by nonresident workers who have lived in the host country for less than a year. According to these data, in 2005 the United States was by far the leading host country for temporary foreign workers, followed by Saudi Arabia, Switzerland, Germany, Spain, the Russian Federation, Italy, Malaysia, the Netherlands, and France, in that order.

Significance for the World Economy The economic significance of TMNP for the world economy stems from the fact that the majority of temporary workers move from areas of low productivity to areas of higher productivity, thus bringing about global efficiency gains. Projections such as those made by Winters et al. (2003) also predict large economic gains for the world economy if TMNP is liberalized. Such studies also usually predict that the gains will be shared equally between sending and receiving countries. Despite this, both developed and developing countries have adopted a conservative policy approach with regard to the entry of foreign workers compared with other areas of international trade. WTO commitments on TMNP attached to the Third Protocol to the GATS are mostly confined to highly skilled labor and typically qualified by restrictions and exemptions. Part of the reason is the recognition that the labor market consequences of trade liberalization unravel differently for different sectors. For instance, the impact on the labor market of lower tariffs on industrial products occurs indirectly, and domestic workers lose jobs after a time lag. Liberalizing the entry of foreign workers, on the other hand, creates immediate and direct competition for domestic workers and is therefore much more challenging politically.

Experts have compiled lists of benefits and drawbacks of permanent migration for both source and host countries. The economic consequences of TMNP generally are the same as those of permanent migration, with a few important exceptions:

- There is no permanent brain drain caused by TMNP, although there is temporary loss of the migrant's contribution to home country output. It is argued that a returning migrant can bring back new ideas, technologies, and networks that would contribute to improved efficiency and productivity and partially compensate for the loss to national welfare caused by temporary absence.
- With TMNP, the option to post and rotate senior corporate personnel of choice abroad improves efficiency and control in the international operations of source country firms. This is especially helpful if qualified local personnel are not available in sufficient numbers in the host country.
- Unlike permanent migrants, temporary workers are less likely to put pressure on host-country public goods or infrastructure or undermine a foreign country's culture, customs, and values on account of the transitory nature of stay abroad.

- Remittances are likely to be higher from temporary migrants, so host countries are likely to experience a higher outflow of earnings under TMNP compared with permanent migration.

See also General Agreement on Trade in Services (GATS); migration, international

FURTHER READING

Bhatnagar, Pradip. 2004. "Liberalizing the Movement of Natural Persons A Lost Decade?" *The World Economy* 27 (3): 459 72. Covers the various economic and noneconomic arguments for and against the liberaliza tion of the temporary movement of natural persons.

Mattoo, Aaditya, and A. Carzaniga, eds. 2003. *Moving People to Deliver Services.* Washington, DC: World Bank and Oxford University Press. Compiles viewpoints of economists, regulators, negotiators, and private sector stakeholders on the subject.

Nielson, Julia. 2003. "Labor Mobility in Regional Trade Agreements." In *Moving People to Deliver Services,* edited by Aaditya Mattoo and Antonia Carzaniga. Wash ington, DC: World Bank, 93 112. An excellent source of information on TMNP in different regional trading arrangements.

Winters, L.A., T. L. Walmsley, Z. K Wang, and R. Gryn berg. 2003. "Liberalizing the Temporary Movement of Natural Persons: An Agenda for the Development Round." *The World Economy* 26 (8): 1137 61. One of the more recent in a series of endeavors to measure the global economic gains of liberalizing TMNP.

World Bank. 2007. *World Development Indicators.* A quick guide to leading source and host countries for TMNP using indirect data like remittances and wages and sal aries earned by nonresidents. Released annually in CD ROM format.

World Trade Organization. 1998. *Presence of Natural Per sons (Mode 4). Background Note by the Secretariat (S/C/W/ 75).* Contains the main definitions, explanations, and a summary of GATS commitments made on the move ment of natural persons in the GATS. Download able from http:/www.wto.org/english/tratop e/serv e/ w75.doc

PRADIP BHATNAGAR

■ terms of trade

The current world economic situation is the legacy of the days of industrial revolution in the North and colonization of the major countries of the South. This divided the world economy into two regions: the industrially developed North and less-developed South. In the context of the North-South divide of the world economy, the long-term behavior of the terms of trade between the two regions has been a much discussed and much debated topic.

By terms of trade we usually mean commodity terms of trade, the term popularized by Jacob Viner. The commodity terms of trade of the South in rela-tion to the North can be defined as the price (unit value index) of exports of the South to the North divided by the price (unit value index) of exports of the North to the South.

The commodity terms of trade is also known as net barter terms of trade, thanks to Frank William Taussig. He introduced another concept, the gross barter terms of trade. It is the ratio of the volume index of imports to the volume index of exports. It is not widely known or used in the literature.

G. S. Dorrence introduced the concept of the income terms of trade. It is an index of the value (= volume x unit value) of exports divided by the unit value of imports and it corresponds to the com-modity terms of trade (= unit value of exports/unit value of imports) multiplied by the volume of ex-ports. It measures the purchasing power of exports the amount of imports that can be financed by total exports of a country or region.

If, instead of volume of exports, the commodity terms-of-trade index is multiplied by some index of factor productivity in the export sector, we get the single factorial terms of trade. To construct the double factorial terms of trade, the single factorial terms of trade is deflated by the index of factor pro-ductivity in the export sector of its trading partner. So, the double factorial terms of trade is the com-modity terms of trade multiplied by the ratio of the indices of factor productivity in the export sec-tors of the trading partners. Usually labor produc-tivity is considered for factorial terms-of-trade index calculation.

The single factorial terms-of-trade index has a simple intuitive implication: it indicates the amount of wheat (say) that can be imported by a country (say, Ghana) by employing one hour of labor in production of its export item, say, cocoa. An improvement in the index implies a rise in the country's labor productivity in the export sector valued in terms of its import goods. Among other possibilities, it may be due to a rise in the labor productivity in its export sector not fully counterbalanced by a (resulting) fall in its export prices relative to its import prices (that is to say, a fall in its commodity terms of trade)   trade has not wiped out the whole fruit of technical progress in its export sector. In the process, its trading partner can experience a rise in its single factorial terms of trade through an improvement in its commodity terms of trade.

There are many other cases where the single factorial terms of trade of two trading countries can rise or fall together, but a rise (fall) in the double factorial terms of trade of one country must imply a fall (rise) in the double factorial terms of trade of its trading partner. That is why the economists interested in relative gains from trade or unequal exchange focused on this concept of terms of trade.

Suppose the North and the South experience the same rate of technical progress and labor productivity improvement in their respective export sectors. One region (say South) experiences lower export prices in response to technical progress and cost reduction   wages do not rise in proportion to the rise in productivity because of (say) surplus labor and the consequent weak labor union. Suppose the other region (North) experiences rising wages and profit because of strong labor union and monopoly power of the producers and no fall in their export prices. Then double factorial terms of trade would move in favor of the North and against the South. This is what was highlighted by Prebisch (1950) and Singer (1950) in the context of trade between the industrially developed North and the less-developed South and gave birth to the well-known Prebisch-Singer hypothesis. They provided some evidence of a long-term decline in the commodity terms of trade of primary products and the primary-product exporting South since the last quarter of the 19th century and indicated a long-term decline in the double factorial terms of trade of the South vis-à-vis the North on the basis of the well-accepted fact that the industrial sector of the North experienced a higher rate of technical progress in those days.

This Prebisch-Singer hypothesis generated much controversy in the academic world both on empirical and theoretical grounds, and it was virtually discarded in mainstream economics. But there is now an increasing volume of literature in its support (see Sarkar 1986, 2001).

In the "mainstream" neoclassical, perfect competitive framework, each factor (including labor) is paid according to its productivity. So the single factorial terms of trade is the real wages measured in terms of import goods and the double factorial terms of trade is the wage ratio (measured in a common currency). In a simple classical political economy framework of the (North-South) world economy with free mobility of capital and labor, the North-South commodity terms of trade would be equal to the ratio of labor embodied in the production of export goods in the two regions (in accordance with the classical labor theory of value) and the double factorial terms of trade would be equal to one. This is the ideal situation   "equal exchange"   the Marxist economist Emmanuel (1969) had in mind. But in a world of nearly perfect mobility of capital and imperfect labor mobility, the commodity terms of trade is not in accordance with the labor theory of value. The double factorial terms of trade are equal to the wage ratio as in the neoclassical framework. Granted that Northern workers get higher wages (even after adjusting for their higher skill and productivity), the double factorial terms of trade of the South are less than one. That implies that one hour of Southern labor is exchanged for less than one unit of (comparable) Northern labor. This is a case of (static) "unequal exchange" at a point of time. If over time this North-South wage gap rises, the double factorial terms of trade would decline indicating increasing "unequal exchange." Thus there is a convergence between the Prebisch-Singer hypothesis and Emmanuel's dynamic "unequal exchange" idea.

Last there is another concept, employment-corrected double factorial terms of trade, introduced by Spraos (1983). It is based on the idea that if a country or region such as South has a large pool of unemployed labor force, a decline in its double factorial terms of trade is welcome if that is more than compensated by a rise in employment through the growth of trade and production. This index is calculated by multiplying the double factorial terms of trade of a country or region by the total employment in its traded sector (if both the countries or regions face this unemployment problem the multiplicative factor would be the employment ratio and the index would be turned into a ratio of export values!). On the assumption of near full employment in the North and a large-scale unemployment problem in the South, he calculated such indexes and found some evidence of decline in the terms of trade of the South from 1960 to 1977.

The concepts of terms of trade and a study of their behavior are useful in understanding whether the evolution of the world economy led to some kind of uneven development through the terms-of-trade movements against the interest of the less developed South, thereby hampering the process of catching up.

See also import substitution industrialization; North-South trade; primary products trade; trade and economic development, international

FURTHER READING

Emmanuel, Arghiri. 1969. *Unequal Exchange*. New York: Monthly Review Press (English translation, 1972). Develops the concept of unequal exchange on the basis of Marxian concepts and tools of analysis.

Prebisch, Raúl. 1950. *The Economic Development of Latin America and Its Principal Problems*. New York: United Nations. A critique of the Ricardian theory of comparative advantages and the international division of labor based on the comparative advantage pattern, this pamphlet also argues in favor of industrialization in the less developed countries to avoid their terms of trade loss.

Sarkar, Prabirjit. 1986. "The Singer Prebisch Hypothesis: A Statistical Evaluation." *Cambridge Journal of Economics* 10: 355 71. Questions the validity of the major statistical objections against the Prebisch Singer hypothesis.

. 2001. "The North South Terms of Trade: A Re examination." *Progress in Development Studies* 1 (4): 309 27. Surveys the literature concerning the terms of trade debate and discusses the theoretical issues.

Singer, Hans. 1950. "The Distribution of Gains between Investing and Borrowing Countries." *American Economic Review* 40: 473 85. Presented at the American Economic Association annual meeting (May 1949), this paper discounts the benefits of foreign investment and technical progress in the primary product exporting sector of the less developed countries because of the consequent terms of trade loss.

Spraos, John. 1983. *Inequalizing Trade?* London: Clarendon Press. Discusses different concepts of terms of trade, their welfare implications, and other issues between the North and the South.

PRABIRJIT SARKAR

■ **textiles and clothing**

The textiles and clothing supply chain consists of production of raw materials; spinning, weaving, or knitting; finishing; design; sewing; distribution; and marketing. The production of textiles was the first economic activity to be industrialized, starting in the United Kingdom around 1765 with the invention of the spinning jenny, a machine for spinning wool or cotton.

The clothing industry is unskilled and labor intensive. The technology is cheap and simple and did not change much over the 20th century. The fabric is first cut and then grouped, tied into bundles (pre-assembly), and sewn together. A worker receives a bundle of unfinished garments, performs her single task, and places the bundle in a buffer. It takes about 40 operations to complete a pair of pants, and about 100 operations for men's blazers. Improvements in technology have taken place in two areas: specialized machines for each task and better coordination of the tasks.

The textile and clothing sector also includes market segments where innovation, technology, and design are important for competitiveness. Examples are sophisticated textile materials for industrial and

medical use, athletic sportswear and equipment, and up-market fashion clothing. In the United States about two-thirds of textiles production is for industrial use and household goods. Italy and, to a lesser extent, France are important producers and exporters of up-market fashion products, and the two combined accounted for about 10 percent of world clothing exports in 2005. In Italy, textiles and clothing as a share of total manufacturing value added declined from 12 percent of total manufacturing value added in 1980 to 10 percent in 2003.

Economies of scale in distribution favor large supplier countries that cover the entire supply chain, such as China, India, Mexico, and Turkey. Smaller producers tend to find niches where they establish a significant market position. Sri Lanka, for instance, is among the world's leading exporters of lingerie.

Although production technology has not fundamentally changed, distribution technology and the division of labor between manufacturers and retailers have changed substantially since the late 1970s. The shift of key functions from manufacturers to retailers started with the establishment of large discount retailers such as Wal-Mart in the United States. Using modern information technology to improve supply chain management and transforming the relationship between retailers and manufacturers, Wal-Mart pioneered a practice that has been coined "lean retailing."

Lean retailers insisted that suppliers take up information technologies in order to facilitate more efficient supply chain management. They typically require manufacturers to adopt their standards and make apparel shelf-ready, which includes price tags and other required labeling. This ensures quick replenishment of apparel, which in turn allows the retailer to offer a broad variety of clothing without holding much inventory. The attractiveness of such contracts for suppliers is first and foremost the scale of operation of the lean retailers, which provides a stable marketing outlet and often also low-cost fabric and other material inputs purchased in bulk by the retailer.

With the introduction of private labels or store brands, design has shifted from the manufacturer to

retailers. Specialized clothing retailers such as H&M and The Gap led this development. Most retailers of cheap, mass-market apparel source private label/store brand items from manufacturers in low-cost countries. However, a much studied Spanish retailer (Zara, which is part of the Inditex Group) has combined lean retailing with vertical integration of design, manufacturing, and distribution. Zara is among the world's largest retailers with retail outlets in 68 countries in 2007. Its innovation is production in small batches, introducing new designs as frequently as twice a week. New designs are based on information from the retail outlets. According to the company's Web page, Zara largely manufactures in Spain, arguing that the benefits of being close to the market outweigh the higher cost of production in Spain compared with low-cost countries. The numerous studies of this retailer suggest that its business model may multiply.

As a labor-intensive industry, textiles and clothing have always been an important source of employment for unskilled workers, particularly women. Furthermore the sector typically clusters geographically, for instance in the Carolinas and Georgia in the United States and in the Emilia-Romagna district in Italy. Because the industry tends to cluster in regions where it constitutes an important source of employment, job losses, for example, due to trade liberalization, can have serious local labor market consequences. Trade in textiles and clothing has therefore always been politically sensitive, and the sector was governed by an international trade regime of import quotas from the early 1960s until 2005.

International Trade Regime The textiles and clothing sector has a long history of import protection. As new countries emerged as important exporters of textiles and clothing, demand for protection arose in the United States and other developed countries. Thus in the 1950s the United States entered into agreements on voluntary export constraints on cotton textiles and clothing with Japan, Hong Kong, India, and Pakistan. In 1962, a Long-Term Agreement Regarding International Trade in Cotton Textiles (LTA) was negotiated under the auspices of the General Agreement on Tariffs and

Trade (GATT). The LTA was extended to materials other than cotton in 1974 and became known as the Multifiber Agreement (MFA).

The MFA governed a trade regime in which the major importing countries negotiated import quotas bilaterally with exporting countries. It became part of the multilateral trading system in spite of breaking fundamental principles of the GATT, including the most-favored-nation (MFN) principle and the idea that trade restrictions should be in the form of tariffs, as tariffs are less trade distorting than quotas. The MFA was renegotiated four times, the last time in 1991, and it finally expired in 1994. At that point the European Union (EU), the United States, Canada, Norway, Austria, and Finland (Austria and Finland joined EU in 1995) applied quotas, almost exclusively on imports from developing countries.

While the most competitive textiles and clothing exporters such as China, India, Pakistan, Hong Kong, Taiwan, and South Korea faced binding quotas, several poorer developing countries had unfilled quotas and therefore attracted foreign investment from the quota-restricted countries. In addition, former colonies and other low-income countries obtained preferential access to the EU and the United States, including tariff and quota-free access in textiles and clothing. This further encouraged the establishment of export-oriented clothing production, often in export processing zones. Preferential access was typically conditioned on a minimum local content (intermediate input) requirement or complicated rules of origin.

Many low-income countries have a narrow industrial base and cannot easily source materials used in the manufacturing of apparel locally and fulfill the local content requirements. In such cases imports of material inputs from other low-income regional trading partners or preference-receiving countries were sometimes allowed. Finally, the United States and the EU had production-sharing programs in which the partner countries could import textiles or cut fabric and other inputs from the United States or EU and export ready-to-wear clothing back to them duty free or paying tariffs only on the value added in the exporting country. The relevant program in the

United States was the 807/9802 production-sharing program (referring to chapter 9802 in the Harmonized Tariff Schedule of the United States relating to articles assembled abroad with components produced in the United States), while the EU had outward processing agreements with several countries, both as part of regional free trade agreements and as stand-alone agreements with a number of Asian countries.

The managed trade system   including the MFA, preferential access to major markets for small, low-income countries, and outward processing agreements   resulted in a highly distorted trade and production pattern. It has been estimated that the MFA regime cost the average U.S. household $105 per year, and the average EU household €270 per year around the year 2000. During the Uruguay Round of the GATT it was agreed to gradually phase out the MFA and subject the textiles and clothing sectors to the normal multilateral trade rules in the GATT. The MFA was replaced by a transitory regime known as the Agreement on Textiles and Clothing (ATC) in 1995. ATC envisaged the gradual phasing out of quotas over 10 years, ending January 1, 2005. Quotas were to be eliminated in four steps in 1995, 1998, 2002, and 2005, while the quotas remaining at each point in time were to be increased every year. In 2004 about 20 percent of U.S. imports came in under a binding quota.

While the ATC was followed to the letter, liberalization was largely back-loaded until 2005, as the nonbinding quotas were lifted first. When China became a member of the WTO in 2001, its considerable textiles and clothing sector became part of the liberalized multilateral trading system. Due to the size and competitiveness of China's manufacturing sector, many studies predicted that China and to a lesser extent India would dominate the post-quota world markets, to the detriment of preference-receiving developing countries and producers in the EU and the United States. Therefore safeguards were incorporated in China's Accession Agreement with the WTO.

When the remaining quotas were lifted on January 1, 2005, there was indeed a surge in imports to the

United States from China. Imports increased by more than 50 percent in value terms from 2004 to 2005, and China's market share increased from 20 to 28 percent in textiles and clothing. This surge triggered requests for safeguards, and the United States signed a Memorandum of Understanding with China in November 2005. The two countries agreed to set annual quantitative levels of imports in 22 categories of textiles and clothing for 2006, 2007, and 2008, in which import expansion volumes of 173 640 percent were allowed between 2004 and 2006 and thereafter by increases of 12.5 16 percent in 2007 and 15 17 percent in 2008.

Imports from China into the EU surged by almost 100 percent in value terms in the liberalized categories during the first three quarters of 2005 and the increase was even sharper in volume terms, as unit prices fell sharply. Imports from China largely replaced exports from other developing countries, while the total import value did not change much. This triggered a Memorandum of Understanding with China, and safeguards on 10 sensitive products were imposed, allowing imports to grow by between 10 and 12.5 percent for the years 2005, 2006, and 2007 in these products. It should also be noted that China has become one of the fastest-growing markets for EU textiles exports.

Both tariffs and quotas impose a wedge between world market prices and the domestic price in the importing country, allowing local high-cost producers to sell in the local market. Quotas have the additional effect that exporters shift to higher-value-added, higher-quality products within the quota in order to maximize income from the allocated quota. A sharp fall in unit prices both in the EU and the United States on imports from China in liberalized product categories during 2005 suggests that this effect may have been important, although elimination of quota rents and more competitive markets have also contributed to lower unit prices.

In spite of import protection, employment in the textiles and clothing sector has declined steadily over time in the United States and the EU. Furthermore the declining trend has continued, but apparently not accelerated, during the phaseout of import quotas under the ATC. In the United States, employment in the textiles sector fell from 701,000 in 1990 to 539,000 in 2001 and further to 390,000 in 2005. The clothing sector experienced an even sharper fall in employment, from 929,000 in 1990 to 426,000 in 2001 and further to 260,000 in 2005. The EU experienced a similarly falling long-term employment trend, from 2.6 million workers in 1995 to 2.2 million in 2001.

Even in the countries with an expanding textiles and clothing sector, employment has increased much less than output as increased international competition has led to the adoption of modern technology and sharp increases in labor productivity. The textile and clothing sectors employed 4.8 and 2.6 million people, respectively, in China in 2002, together about 16.5 percent of China's total manufacturing employment.

World Trade in Textiles and Clothing Textiles and clothing's share of world merchandise exports in value terms has fluctuated between 5 and 8 percent since the early 1960s and stood at 4.7 percent in 2005. The direction of trade has shifted substantially over the decades as textiles and clothing production tends to migrate to countries with relatively low labor costs. The first wave of migration came in the 1950s and 1960s from Europe and the United States to Japan. The second wave came in the 1970s and early 1980s from Japan to the rapidly industrializing countries in Asia: the Republic of Korea, Taiwan, and Hong Kong, which together accounted for about a third of world clothing exports at their peak. Finally there is an ongoing migration of production and trade from these middle-income Asian countries to China, India, Bangladesh, Vietnam, and other South Asian countries.

China has increased its world market share in clothing from 6 percent in the early 1980s to 26 percent in 2005, while India's world market share has fluctuated between 2 and 3 percent since the early 1980s. India's exports have been hampered by quotas and by domestic labor and product market regulation, but are likely to increase sharply in a quota-free trading regime, provided that domestic reforms take hold.
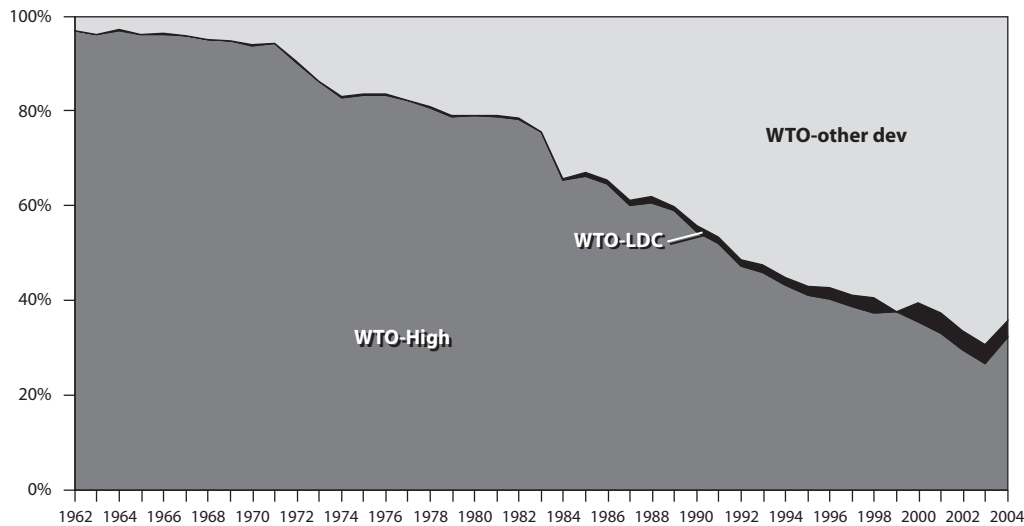
Figure 1

Exports of clothing by income group, shares of total. Source: Comtrade database (UNCTAD/World Bank).

The migration of clothing from high-income countries to developing countries is illustrated by figure 1, which depicts the shares of total WTO member countries' exports for high-income countries, least-developed (poor) countries, and other developing (middle-income) countries during the period 1962 2004. The countries that are currently classified as middle-income countries have increased their world market share from about 3 percent in 1962 to almost 70 percent in 2004. The countries currently classified as least-developed countries had a negligible world market share until the mid-1980s, but have since established themselves in the export markets. In 2005 the 10 largest exporters of clothing were China, the EU (extra-EU trade only), Turkey, India, Mexico, Bangladesh, Indonesia, the United States, Vietnam, and Romania. China alone accounted for 27 percent of total world exports, up from 4 percent in 1980 and 18 percent in 2000, while Hong Kong added another 10 percentage points in 2005 (with domestic exports plus reexports). The 10 largest exporters in textiles in 2005 were the EU (extra-EU trade only), China, the United States, South Korea, Taiwan, India, Pakistan, Turkey, Japan, and Indonesia.

The export success of some least-developed countries is largely based on preferential access to the quota-restricting countries. Some of these countries are not competitive under the normal market conditions that are emerging following the phasing out of quotas. Thus exports of textiles and clothing from ACP countries (African, Caribbean, and Pacific countries with preferential access to the EU market) to the EU fell by 17 percent in 2005.

With the increasing importance of design and rapid response to shifting consumer preferences and fashion trends, there has been a trend toward regionalization of the textiles and clothing trade. This development has been masked by the lifting of quotas where China and India were the most restricted countries. Mexico is the second largest exporter (after China) of clothing to the United States, however, while Turkey is the second (after China) and Romania the fourth (after India) largest exporter of textiles and clothing to the EU.

See also commodity chains; export promotion; fragmentation; quotas; safeguards

FURTHER READING

Abernathy, F. H., J. T. Dunlop, J. H. Hammond, and D. Weil. 1999. *Lean Retailing and the Transformation of*

Manufacturing Lessons from the Textile and Apparel Industries. Oxford: Oxford University Press. Describes the textiles and clothing supply chain, focusing on the role of retailers, and documents how advances in communication technology have transformed the sector from being supply driven to becoming demand driven.

Eurostat. http://epp.eurostat.ec.europa.eu. Source of employment data for the EU.

Evans, C. L., and J. Harrigan. 2005. "Distance, Time, and Specialization: Lean Retailing in General Equilibrium." *American Economic Review* 95: 292 314. Provides empirical evidence that time is an important competitive factor in the clothing sector and finds that the sources of U.S. imports of "fast fashion" items have shifted to nearby countries.

Rivoli, P. 2005. *The Travels of a T shirt in the Global Economy: An Economist Examines the Markets, Power, and Politics of World Trade.* Hoboken, NJ: John Wiley and Sons. Follows a T shirt from the cotton fields in Texas to factories in China, sales in the United States, and resale on the secondhand clothing market in Dar es Salaam. At each leg of the journey, economic, policy, and social aspects are discussed.

UN Comtrade. http://comtrade.un.org/db/. Source of trade data.

UNIDO. http://www.unido.org/doc/3474. Source of employment data for countries other than the United States and the EU.

U.S. Bureau of Labor Statistics. http://www.bls.gov/ces/home.htm. Source of employment data for the United States.

World Trade Organization (WTO). 2001. "Comprehensive Report of the Textiles Monitoring Body to the Council of Trade in Goods on the Implementation of the Agreement on Textiles and Clothing during the Second Stage of the Integration Process." G/L/459 (July), 31. An official WTO report from the Textiles Monitoring Body assessing the implementation of the ATC, it provides comprehensive documentation of the implementation of the Agreement on Textiles and Clothing and its market access impact and can be downloaded from the WTO Web site.

HILDEGUNN KYVIK NORDÅS

■ time inconsistency problem

Suppose a government is responsible for setting a policy for several periods into the future. If the government chooses to change its policies from those promised at an earlier date, the policies are said to be *time inconsistent*. The modern interpretation of this issue relates to the time inconsistency of optimal policy rules (Kydland and Prescott 1977). Policymakers may announce in advance the policies they will follow to influence the expectations of private households. Once private decision-makers have acted on the basis of their expectations, however, a policymaker may be tempted to renege on an announcement, leading to suboptimal outcomes.

For example, consider the example of a central bank that is concerned about the inflation rate and the unemployment rate in an economy. The Phillips curve, which postulates an inverse relationship between inflation and unemployment, predicts that the trade-off between inflation and unemployment depends on expected inflation (Mankiw 2007). If policy setting in the central bank is guided by the Phillips curve, to reduce expected inflation the central bank might announce that low inflation is the main goal of its monetary policy. This is because a favorable trade-off between unemployment and inflation depends on the expected inflation rate being low. The announcement of a policy of low inflation is not credible, however. The central bank has an incentive to renege on its announcement and implement an expansionary monetary policy to further reduce the unemployment rate once firms and households have formed their expectations of inflation for the future, and set wages and prices accordingly. Private decision-makers understand the central bank's inherent incentive to shirk its announced policy and do not believe the central bank's announcement in the first place. The implication is that the economy ends up with a higher inflation rate without any lowering of the unemployment rate. Time inconsistency therefore leads to suboptimal outcomes.

Insights from the time inconsistency problem underlie the widely accepted contemporary view among economists that the best policies are the most

predictable policies, or those that follow simple rules. This basic insight also relates to an earlier debate on rules versus discretion in macroeconomic policy (Friedman 1959). A rule specifies in advance the actions that policymakers will take or commit to. Commitment refers to the ability of a government to make binding choices. Economists have long recognized that unless a government can commit to a policy rule, it will in the future want to modify policies that are optimal from the standpoint of today. Further, without such binding policy commitments (i.e., policies aimed at tying the hands of the government), the private sector will fear that today's governments will make promises that its successors will refuse to honor.

Monetary Policies to Overcome Time Inconsistency Time inconsistency provides a rich explanation for a wide variety of issues, such as why capital accumulation in countries is low, or why countries choose to fix their exchange rate. In line with the earlier example, however, one of the most important applications of time inconsistent policies is the "inflation bias" inherent in monetary policy (Barro and Gordon 1983a, 1983b). *Inflation bias* refers to prolonged episodes of high inflation above what is socially desirable. An example is the "great inflation" of the 1970s in the United States. Time inconsistency hampers a government's efforts to keep inflation stable and leads to high and persistent inflation despite repeated promises to fight it. This affects the volatility of inflation, output, and interest rates, leading to a misallocation of resources.

Given the implications of time inconsistency, a large body of research has evolved since the 1980s on ways to overcome the time inconsistency problem in macroeconomic policy. For example, in the case of monetary policy, economists have suggested various types of monetary policy rules. Other solutions involve the delegation of monetary policy to an independent authority with well-defined objectives. This in turn has led to greater central bank independence. Central bank independence can help in situations where credible policy rules are not enforceable. In contrast to rules,

discretion implies the absence of commitment to a particular rule, leaving policymakers more flexible with their future actions.

Another solution to the time inconsistency problem involves reputation building. By acting consistently over a long period, a policymaker builds up a reputation that causes the private sector to believe its announcements. Once a reputation has been built, expectations of inflation become consistent with the announced low inflation policy. In some cases reputation can be imported. For instance, the monetary authority can peg its currency and import the monetary policy of another country with more credible institutions. This is often the primary motivation behind the adoption of hard pegs such as currency boards, dollarization, and common currencies.

Political Macroeconomics Since the early 1990s a line of research called political macroeconomics has evolved to explore the political determinants of macroeconomic policy (Drazen 2000; Persson and Tabellini 1994). This line of research recognizes that a policymaker's incentives and constraints originate in the political process. For example, within the context of fiscal policy, elected officials may be motivated by electoral reasons to enact time inconsistent expansionary policies that have long-term inflationary consequences. Alternatively, fiscal stabilizations attempts to reduce a large fiscal deficit of a country are likely to fail if the adjustments are achieved by higher taxes, as in Ireland's fiscal stabilization of 1982. The time inconsistency of this policy is that governments are unlikely to deliver on the promise of higher taxes, given the nature of the political business cycle. With respect to monetary policy, the economist Rogoff (1985) argues for the appointment of a conservative central banker who has greater inflation aversion than society as a whole to eliminate the inflation bias inherent in monetary policy. Welfare improves because the central bank trades some flexibility for some gain in credibility. Indeed, different views about rules and discretion with the associated themes of commitment, credibility, and time inconsistency are

among the most pressing issues concerning the conduct of macroeconomic policy in a strategic setting.

See also common currency; currency board arrangement (CBA); currency substitution and dollarization; discipline; exchange rate regimes; inflation targeting; monetary policy rules; monetary versus fiscal dominance

FURTHER READING

Barro, R., and D. Gordon. 1983a. "Rules, Discretion, and Reputation in a Model of Monetary Policy." *Journal of Monetary Economics* 12: 101 21.

———. 1983b. "A Positive Theory of Monetary Policy in a Natural Rate Model." *Journal of Political Economy* 91: 589 610. These two articles show how an incentive toward time inconsistent behavior in the Kydland Prescott model can yield an inflationary bias in monetary policy. They also introduce a new solution to the time inconsistency problem: reputation building.

Drazen, A. 2000. *Political Economy in Macroeconomics.* Princeton, NJ: Princeton University Press. Represents the first full scale effort to organize a large body of literature on political economy and macroeconomics. It has an extensive coverage and depth as well as critical assessment.

Friedman, M. 1959. *A Program for Monetary Stability.* New York: Fordham University Press. Contains a comprehensive discussion of an early debate between rules versus discretion in macroeconomic policy.

Kydland, F., and E. Prescott. 1977. "Rules Rather than Discretion: The Inconsistency of Optimal Plans." *Journal of Political Economy* 85: 473 91. The modern interpretation of time inconsistency of policy stems from this pathbreaking paper. It had a profound effect on the practice of policymaking.

Mankiw, G. 2007. *Macroeconomics.* 6th ed. New York: Worth. Contains economic and noneconomic examples on time inconsistency.

Persson, T., and G. Tabellini. 1994. *Monetary and Fiscal Policy.* Vol. 1. *Credibility.* Cambridge, MA: MIT Press. Contains a full treatment of political macroeconomics with a specific focus on policy issues.

Rogoff, K. 1985. "The Optimal Degree of Commitment to an Intermediate Monetary Target." *Quarterly Journal of Economics* 100: 1169 90. A pathbreaking paper in the time inconsistency literature on the importance of central bank independence.

CHETAN GHATE

■ Tinbergen principle/rule
See assignment problem

■ Tobin tax

Nobel Laureate economist James Tobin (1974, 1978) advanced a now well-known case for what has come to be known as the Tobin tax, namely, the imposition of a modest ad valorem tax (i.e., a percentage of value) on all spot transactions in foreign exchange. (Spot transactions are those that entail actual exchanges of currency at the existing market rates.) Tobin wrote that "the tax would apply to all purchases of financial instruments denominated in another currency. It would have to apply . . . to all payments in one currency for goods, services, and real assets sold by a resident of another currency area" (1978, 159). Tobin advanced the case for this tax on currency trading following the collapse of the Bretton Woods system of pegged exchange rates and attendant concerns about speculation, volatility, and misalignment in currency markets. Tobin also suggested that the tax could restore "some fraction of short-run [policy] autonomy"; however, he was careful to explain that "it will not, should not, permit governments to make domestic policies without reference to external consequences" (1978, 158).

The Tobin tax built on John Maynard Keynes's case for a securities transactions tax. Keynes (1936) proposed a substantial transfer tax on all transactions as a means to dampen the general tendency for speculation to dominate enterprise in liquid, competitive financial markets. Keynes (1980, chap. 36) also proposed taxation of foreign lending to contain speculative capital movements.

Several factors stimulated interest in the Tobin tax in the 1990s. These factors included the increase in currency speculation and volatility; the frequency of currency crises in developing countries; and the debate over the adequacy of the international financial architecture and the volume of international capital flows available to developing countries. This attention stood in sharp contrast to the silence that greeted Tobin's original presentation in the 1970s. The recent interest in the Tobin tax is located in a number of quarters namely, academic economists identified with the Keynesian tradition, various United Nations agencies, governments in some countries that express support for the Tobin tax (namely, Canada, Belgium, and France), and nongovernmental organizations concerned with development finance and financial architecture. Tobin distanced himself from many of the nongovernmental organizations that advocated for the tax in the 1990s, however.

There is broad consensus that the tax must be levied at a low rate (Nissanke 2004 suggests 0.01 to 0.02 percent) in order to minimize the incentive to undertake tax evasion strategies (e.g., geographical or asset substitution) and to avoid other distortions of the foreign exchange market. Research has also focused on the issue of tax collection. In Tobin's original formulation, governments would levy and collect the tax nationally. But in view of the potential problems of tax competition among nations and the diversion of trading activity to "Tobin tax havens," many proponents focus on the need for a global tax agreement and the creation of a supranational authority to administer and collect the tax (Griffith-Jones 1996) and to allocate its revenues to projects that promote development (Kaul and Langmore 1996). Others, however, suggest that a Tobin tax could achieve some of its objectives absent global implementation (Baker 2001; Nissanke 2004).

Modifications to the Original Tobin Tax Concept In 1996 Tobin amended his original proposal to take account of the importance of new instruments of foreign exchange trading. In the amended version, Tobin (1996) argued that forward and swap transactions would be subject to taxation as well. (A for-

ward transaction is a contract in which parties agree to exchange currencies in the future at a price agreed on today. A swap combines spot and forward transactions.) Indeed, it is likely that if the tax is limited to spot transactions it would lead to a tax-saving reallocation of financial transactions from traditional spot transactions to derivative instruments. As such, in order to prevent tax avoidance via asset substitution or a changed product mix, it ought to be applied on all derivative products such as forwards, futures, options, and swaps.

Some proponents of the Tobin tax argue for modified versions of the proposal aimed at enhancing its ability to reduce currency market volatility. For instance, Crotty and Epstein (1996) argue for joint implementation of Tobin and securities transaction taxes. They argue that a securities transaction tax might reinforce the market-stabilizing function of a Tobin tax by increasing the cost of investor flight (since the sale of assets or borrowing must precede the flight of large sums of money from a country).

Spahn (1996) advances a case for a two-tiered Tobin tax. In this formulation, low transactions taxes on currency trading would be maintained during tranquil (or "normal") times. But a higher transaction tax would be activated whenever levels of market activity accelerated dramatically. Spahn argues that with knowledge of this variable tax structure, investors might be less likely ex ante to engage in activities that aggravate various types of financial risks. Moreover activation of a prohibitively high transaction tax (as a type of speed bump) might discourage some investors from liquidating their portfolios at precisely the time of greatest financial volatility.

The Debate over the Tobin Tax Most proponents of the Tobin tax argue that it has the potential to shift the balance of factors influencing the exchange rate away from short-term expectations toward long-term factors and to decrease exchange rate volatility caused by speculation in this market. Advocates maintain that speculation in currency markets would be dampened because the annualized cost of even a low tax could change trading behavior. Many proponents argue that the Tobin tax consti-

tutes a progressive form of taxation, given its likely incidence (Baker 2001; Palley 2001). Following the East Asian financial crisis of 1997 98, a few advocates suggested that had a Tobin tax been in place it might have reduced the buildup risks that culminated in the crisis (Wade 1998).

Advocates of the Tobin tax are sanguine on the matter of revenue creation. For instance, Nissanke (2004) forecasts that a Tobin tax (even if not implemented by all countries at once) has the potential to raise between U.S. $17 billion and $35 billion in one year (using data for 2001). The taxes harvested would be allocated to various projects of developmental or global importance.

Critics of the Tobin tax, such as Dodd (2003), advance a number of arguments. Dodd argues that the tax is not politically feasible; that it is not feasible administratively or technically without imposing unreasonably high costs; that the issue of leakages through shifts to nontaxed assets and/or to Tobin tax havens is not considered seriously; that the tax would have perverse consequences insofar as it would reduce financial market stability and increase volatility in prices and capital flows; and that the proposed rate for the Tobin tax is actually quite large when one compares it with the transaction costs of trading in foreign exchange markets and most liquid securities and derivatives.

Critics have also taken issue with the claim that the Tobin tax could play any role in reducing the tendency toward financial crises in developing countries (Grabel 2003). This is because the Tobin tax is not designed to dampen speculation in all of the sectors of the economy that are prone to bubbles (e.g., real estate and construction). Even in those sectors that do fall under the authority of the tax, the presence of a tax is unlikely to undermine the attractiveness of activities and financing strategies that aggravate fragile financial environments. Implicit in this critique of the Tobin tax is the view that other strategies ranging from prudential financial regulation, exchange controls, capital controls, capital management techniques, and taxation of other "global public ills" (such as pollution) represent far better means than does the Tobin tax for reducing

financial volatility and currency misalignments, enhancing macroeconomic policy autonomy, and raising revenues for projects of global social importance.

See also Bretton Woods system; capital controls; capital flight; currency crisis; exchange rate volatility; financial crisis; global public goods; international financial architecture

FURTHER READING

Baker, Dean. 2001. "Why Do We Avoid Financial Trans actions Taxes?" *Challenge* 44 (3): 90 96. Focuses on the politics of the Tobin and other types of financial trans actions taxes and argues that the primary obstacle to serious consideration of these measures is political, rather than technical or economic.

Crotty, J., and G. Epstein. 1996. "In Defence of Capital Controls." In *Are There Alternatives? Socialist Register 1996*, edited by Leo Panitch. London: Merlin Press, 118 49. Examines the economic and political argu ments for various types of capital controls, including the Keynes and the Tobin tax.

Dodd, Randall. 2003. "Lessons for Tobin Tax Advocates: The Politics of Policy and the Economics of Market Micro Structure." In *Debating the Tobin Tax*, edited by James Weaver, Randall Dodd, and Jamie Baker. Washington, DC: New Rules for Global Finance Coa lition, 27 50. Presents a critical microeconomic and political analysis of the Tobin tax, takes issue with a great many of the arguments advanced by proponents, and focuses on the unanticipated negative economic and political consequences that would be associated with imposition of the tax.

Grabel, Ilene. 2003. "Averting Crisis: Assessing Measures to Manage Financial Integration in Emerging Economies." *Cambridge Journal of Economics* 27 (3): 317 36. Ex amines the ability of many policies (including the Tobin tax) to reduce financial instability in developing coun tries and concludes that the Tobin tax cannot reduce the most important types of financial risks confronting de veloping countries and, therefore, cannot reduce their vulnerability to crises.

Griffith Jones, Stephany. 1996. "Institutional Arrange ments for a Tax on International Currency Trans actions." In *The Tobin Tax: Coping with Financial*

Volatility, edited by Mahbub ul Haq, Inge Kaul, and Isabelle Grunberg. New York: Oxford University Press, 143 61. Makes a case for the role of a supra national authority charged with administering and collecting Tobin tax revenues.

Kaul, Inge, and John Langmore. 1996. "Potential Uses of the Revenue from a Tobin Tax." In *The Tobin Tax: Coping with Financial Volatility*, edited by Mahbub ul Haq et al. New York: Oxford University Press, 255 72. Argues that Tobin tax revenues should be allocated to projects that promote economic development.

Keynes, John Maynard. 1936. *The General Theory of Em ployment, Interest, and Money*. London: Macmillan. Contains a brief discussion of the need to tax securities trading as a tool for reducing what Keynes saw as the tendency of speculation to dominate enterprise in liquid, competitive markets.

. 1980. *The Collected Writings of John Maynard Keynes*. Vol. 6. *A Treatise on Money*. London: Macmillan. A brief passage in this volume contains the argument that foreign lending should be subject to taxation in order to reduce speculative capital movements.

Nissanke, Machiko. 2004. "Revenue Potential of the Tobin Tax for Development Finance." In *New Sources of De velopment Finance*, edited by A. B. Atkinson. New York: Oxford University Press, 58 90. Presents an estimate of the revenue raising potential of the Tobin tax.

Palley, Thomas. 2001. "Destabilizing Speculation and the Case for an International Currency Transactions Tax." *Challenge* 44 (3): 70 89. Makes a case for the ability of the Tobin tax to reduce speculative pressures and pres ents a number of arguments from the public finance literature in support of the tax.

Spahn, Paul. 1996. "The Tobin Tax and Exchange Rate Stability." *Finance and Development* 33 (2) (June 24 27): 24 27. Advances an argument for a two tiered Tobin tax.

Tobin, James. 1974. "Prospects for Macro economic Pol icy." In *The New Economics One Decade Older*. Prince ton, NJ: Princeton University Press.

. 1978. "A Proposal for Monetary Reform." *Eastern Economics Journal* 4: 153 59. In the 1974 and 1978 papers, Tobin articulates the case for a tax on currency trading, which later became known as the Tobin tax.

. 1996. "Prologue." In *The Tobin Tax: Coping with Financial Volatility*, edited by Mahbub ul Haq et al. New York: Oxford University Press, ix xviii. In this short essay, Tobin updates his 1974 and 1978 arguments in light of changes in financial market structure.

Wade, Robert. 1998. "The Asian Debt and Development Crisis of 1997 ?: Causes and Consequences." *World Development* 26 (8): 1535 53. Examines the causes of the East Asian financial crisis of 1997 98 and offers a brief counterfactual statement to the effect that a Tobin tax might have reduced some of the financial risks that contributed to the crisis.

ILENE GRABEL

■ trade and economic development, international

Since the study of economic development covers a wide array of topics, this entry looks at how international trade affects only certain very important aspects of development, particularly per capita real incomes and their growth rates, and the distribution of income. In studying the impact of trade on distribution-related aspects, the focus is first on poverty, then on the welfare of workers.

The Impact of Trade on Income Levels Theoretically trade is expected to have a positive impact on per capita real incomes through efficiency gains from specialization and exchange, as well as through the availability of larger varieties of goods. Trade has a further beneficial effect on resource allocation through the destruction of the monopoly power of inefficient domestic firms. Recent calculations using standard trade models for a small country show that the gains from moving from no trade to frictionless trade can result in substantial increases in real gross domestic product (GDP). Positive effects of trade on welfare (real incomes), however, may not be obtained in the presence of market distortions, externalities, or imperfections in institutions. Nevertheless, when direct policies to attack these distortions are in place, the positive effects of trade on welfare are restored (Bhagwati 1971).

At the macro level, Frankel and Romer (1999) find a positive effect of trade openness on income levels across countries. They address the issue of reverse causation through an instrumental-variables approach. Rodrik, Subramanian, and Trebbi (2004) use a simultaneous, structural-equations approach to study the effects of institutions, geography, and trade on per capita income levels. They find that "the quality of institutions trumps everything else." Trade and institutions have positive effects on each other, however, so trade affects incomes through institutions.

It is important to mention a few microlevel studies that look at a couple of channels mentioned above. Levinsohn (1993), Harrison (1994), and Krishna and Mitra (1998), using plant/firm-level data have found that the price-to-marginal-cost markup fell as a result of trade reforms in Turkey, Ivory Coast, and India, respectively. Thus the monopoly power of domestic firms (and therefore the deadweight losses associated with this monopoly power) went down. There is also some evidence on the variety channel. Using 8 10-digit level data, Broda and Weinstein (2006) show that the gains from trade for the United States from greater varieties of goods through trade amounts to about 3 percent of GDP. These authors, in their more recent research, extend this empirical analysis and approach to investigating the growth effects of trade through an increase in varieties and find the gains to be even larger. The growth effects of trade will be discussed in detail in the next section.

In summary, the clear theoretical predictions of the positive effect of trade openness on real income or welfare have been supported by macro, cross-country studies. The microlevel studies have identified the main channels to be the market disciplining function of trade and its role in providing greater varieties of goods.

The Impact of Trade on Economic Growth The dynamic effects of trade are not straightforward to analyze and are sensitive to changes in assumptions. As a result, unlike the level effects of trade, there is disagreement on its growth effects.

In the beginning of the second half of the 20th century, the newly independent countries were concerned about possible adverse effects of trade on their growth, as their comparative advantage was in agricultural products. Furthermore they were concerned about adverse changes in the world relative prices of primary products. The early literature on trade and growth is motivated by such concerns.

According to Findlay (1984), the first implicit dynamic model of trade and growth is in David Ricardo's (1815) *Essay on the Influence of a Low Price of Corn upon the Profits of Stock*. This is explicitly and elegantly modeled by Findlay (1984) himself in a two-sector setup. While the profits in manufacturing are invested in increasing production and employment in the next period, land rents in agriculture are spent on consumption by "profligate landlords." Thus a country with a comparative advantage in manufacturing (agriculture) will experience an increase (decrease) in its growth rate on opening to trade.

Similar models of trade and growth were written to study how the North-South terms of trade move over time (Findlay 1984). Feeding stylized facts about relative productivity growth rates in the production of food, coffee, and steel into such models, the implication derived was that terms of trade for the South will deteriorate over time. Working with such models also showed that high growth in the presence of trade could result in a terms-of-trade loss, that is, the good that this rapidly growing country sells (exports) to the rest of the world becomes relatively cheaper while the good that this country buys (imports) from the rest of the world becomes relatively more expensive. When this terms-of-trade loss is large enough, there would be the possibility of a country's own growth reducing its real income. This possibility is called "immiserizing growth," analyzed in detail in Bhagwati (1958).

Another issue studied in the trade and growth literature was the impact of trade on growth through capital accumulation. In neoclassical models of trade and growth, there is either exogenous or no technological progress. Findlay (1984) presents such two-sector Solow- and Ramsey-type growth models to study the effects of trade. A country with a high

savings rate has a high steady-state capital stock in autarky. Such a country will have a comparative advantage in the more capital-intensive of its two sectors. Opening to trade will result in (incomplete) specialization in the capital-intensive sector and will in turn lead to an increase in the rate of return on capital. Incentives to capital accumulation will therefore increase, and we will see an increase in the steady-state capital stock as a result of trade. Opening to trade also leads to transitional growth in this country. A country with a low savings rate has exactly the opposite experience. Importantly, in such multisectoral growth models, unlike a closed economy a small open (but diversified) economy does not experience diminishing returns to capital. With an increase in capital stock, production in a small open economy moves to more capital-intensive goods at a fixed world-price vector.

The issue of capital accumulation in an open economy has also been addressed by the "two-gap" models that emphasize the role of savings and foreign exchange in facilitating growth (Findlay 1984). Investment results in growth but it has to be financed by savings. Since this investment also requires the purchase of foreign capital goods, adequate foreign exchange should be available. That is why these models emphasize the need for aid and promotion of exports in generating growth.

In addition to these types of models, the early trade and growth literature also consisted of North-South models and the dual-economy models. While the former type focuses on the unequal relationship between the developed and the developing world, the latter looks at the asymmetry between regions within the same country. In North-South models, the developing world (the South) is structurally quite different from the developed world (the North) not just in the model parameters but also with respect to the nature of factor markets and institutions. In dual-economy models, there is a modern urban sector and a primitive rural sector. In addition, there could be an unorganized subsector within the urban sector (see Findlay 1984).

While neoclassical growth theory assumed exogenous technological progress, in the endogenous growth theory, pioneered by Romer (1986), the rate of technological progress was determined within the model. Endogenous growth was a major breakthrough in growth research and attempted to explain the growth miracles of East Asia.

Grossman and Helpman (1991) developed a series of models using this new approach, where they clearly show that the growth effects of trade are very sensitive to the model structure assumed. While trade can increase productivity growth by avoiding duplication of effort in research and development (R&D), through the international exchange and transmission of technical knowledge and by allowing the pooling of knowledge across borders, it can reduce or increase productivity growth depending on whether a country is scarce or abundant in skilled labor (Grossman and Helpman 1991).

At the firm level, Rodrik (1992) shows that the market size of a domestic, import-competing firm may shrink as a result of reducing import protection (i.e., allowing more imports) and so will reduce the returns from and investment in R&D and in turn will reduce productivity growth. This is called the "market-size effect"; there is also what Devarajan and Rodrik (1991) call the "procompetitive effect" of trade liberalization, which results in the flattening of the demand for an import-competing good. The greater foreign competition increases the return to R&D and thus also R&D and growth.

At the macro level, the effects of trade barriers on growth and income have been studied empirically since the early 1990s. A variety of openness measures have been used to show positive effects of trade on growth. These studies have come in for heavy criticism by Rodriguez and Rodrik (2001), however, for problems with their openness measures and econometric analysis, as well as for the difficulty in establishing the direction of causality. One of the openness measures criticized by Rodriguez and Rodrik (2001) is the one used by Sachs and Warner (1995), which captures many aspects of the macroeconomic environment in addition to trade policy. However, the Sachs-Warner approach has recently been defended on the grounds that the other policy reforms captured in the measure accompany most major trade

reforms and that we would like to know the value of the entire package of trade and accompanying reforms.

We next turn to the microlevel evidence. Harrison (1994) and Krishna and Mitra (1998) estimate a modified version of the traditional growth accounting equation to incorporate imperfect competition and nonconstant returns to scale. Using plant/firm-level data, they find evidence for an increase in productivity growth after trade reforms in the Ivory Coast and India, respectively. Pavcnik (2002), using a more sophisticated method to correct for simultaneity and selection problems, also finds evidence in support of productivity improvements in Chile after their trade reforms.

In summary, the trade and growth literature began by addressing interactions between trade and the asymmetries across developed and developing countries as well as those within developing countries. Subsequently the focus shifted to the interaction between capital accumulation and trade, in many cases analyzed in the presence of exogenous technical progress. This was followed by the trade and endogenous growth literature, motivated by the East Asian growth miracle. Across these three stages of the literature, the trade and growth models do not agree with each other on the predictions they make, as the results are highly sensitive to changes in assumptions. While most of the empirical literature has shown that trade has a positive effect on growth, the macro cross-country studies have come in for considerable criticism.

The Impact of Trade on Poverty While the literature on trade, incomes, and growth has implications for the trade-poverty linkage, there is also a small literature on the direct determinants of poverty rates (see comprehensive surveys by Goldberg and Pavcnik 2007; and Winters, McCulloch, and McKay 2004).

Dollar and Kraay (2002), using cross-country data, find that the growth rates of average incomes of people in the bottom quintile are roughly equal to the growth rates of overall per capita incomes, and the two are strongly correlated. Also policies known to be growth promoting, such as trade openness, macro-

economic stability, moderate government size, financial development, strong property rights, the rule of law, and the like, promote growth in the incomes of the poor as well. In subsequent work, the same authors infer from the data on the post-1980 "globalizing developing economies" that per capita income growth arising from expansion in trade in those countries has led to a sharp decrease in absolute poverty in the past 20 years.

Ravallion (2001) finds in his cross-country study that an increase in per capita income by 1 percent can reduce the proportion of people below the $1-a-day poverty line by about 2.5 percent on average. This varies across countries, depending on initial inequality.

Finally, a paper by Hasan, Quibria, and Kim (2003) argues, using cross-country evidence, that "policies and institutions that support economic freedom are critical for poverty reduction." Economic freedom indicators used by these authors include, among many components of freedom, the freedom to trade with foreigners.

Moving away from cross-country studies to within-country analysis, Topalova (2005) has examined the impact of trade liberalization on district-level poverty in India. She finds that "rural districts where industries more exposed to trade liberalization were concentrated experienced a slower progress in poverty reduction." She finds this effect not to be statistically significant in urban India. A recent study by Hasan, Mitra, and Ural (2007) uses data at the state and region levels for India and in no case do the authors find reductions in trade protection to have worsened poverty. Instead they find that states whose workers are on average more exposed to foreign competition tend to have lower rural, urban, and overall poverty rates (and poverty gaps), and this beneficial effect of greater trade openness is more pronounced in states that have more flexible labor market institutions. They also find that trade liberalization has led to poverty reduction to a greater degree in states more exposed to foreign competition by virtue of their sectoral composition.

Recent work by Guido Porto at the World Bank looks at the trade-poverty link in a more structural

(model-based) way in a series of papers (e.g., 2006). Using this method, he decomposes the effects of trade into the consumption and income effects. Consumption effects take place through changes in goods prices, while income effects take place through changes in factor incomes. This methodology, developed by Porto, helps us understand how households in different income classes are affected by trade policies. Using this approach, Porto finds a beneficial impact of the South American regional trade agreement Mercosur on average and poor households in Argentina but not on rich households.

The literature on trade and poverty has evolved from cross-country studies to microlevel household studies. The state-of-the-art approach is fairly structural and can identify the various channels through which trade affects poverty. The literature, mainly but not entirely, provides evidence in support of the poverty-reducing effects of trade.

The Impact of Trade on Labor in Developing Countries There is, by now, a fairly rich microempirical literature on how labor is affected by trade policy in developing countries. Recent research draws attention to two opposing effects of trade for which there is some evidence: (1) an increase in wages (the Stolper-Samuelson effect in a labor-abundant country), everything else being equal, which reduces the amount of labor demanded, and (2) a reduction in the price-cost markup (due to more competition), which causes producers to increase their demand for labor. Also, there is evidence that imports can destroy or reduce monopoly rents that can be shared by employers and employees and thus lead to a reduction in the firm-specific wage rate and employment. Currie and Harrison (1997) argue that the employment reducing effects of trade reforms in import-competing firms may be partially or fully offset by positive effects brought about by attempts to reduce profit margins and to invest more resources into productivity enhancement to counter new competition. The mutually opposing forces end up offsetting each other as seen in their firm-level evidence from Morocco. It is important here to point out that many of the studies in the literature, including some of the studies described above, find little change in

the structure of employment across industries as a result of trade reforms, but there is evidence that sectors with larger tariff cuts experience larger declines in relative wages (see Goldberg and Pavcnik 2007).

Another important issue related to the trade-labor link is that of trade and wage inequality. Feenstra and Hanson (2003) summarize their research on this issue, published in a series of papers. Their analysis is based on a new type of trade, namely outsourcing through capital mobility. They assume that there is a final good that is produced using different types of intermediate goods, each of which in turn is produced using skilled and unskilled labor and physical capital. Different inputs have different skill intensities. Under free trade and no capital mobility, the Northern country produces inputs that are more skill intensive than those produced in the Southern country. Under a configuration in which the return to capital is lower in the Northern country than in the Southern country, allowing capital mobility leads to a flow of capital from the North to the South. There is a consequent shift in the production of the North's least skill-intensive intermediate inputs to the South (which still are more skill intensive than all the other inputs being produced in the South). Thus the relative demand for skills and consequently wage inequality go up in both countries. Feenstra and Hanson have found support for their theory using U.S. and Mexican data.

Let us next look at the pressure exerted by trade on labor markets through a new channel that has an impact on employment and wage risk and on the bargaining power of workers. This new linkage is the one between openness and the elasticity of labor demand emphasized by Rodrik (1997) and Slaughter (2001). The elasticity of labor demand here refers to the responsiveness of the quantity of labor demanded to its wage rate. Product-market elasticities (the responsiveness of quantity of final output demanded to the price of output) are likely to rise with trade liberalization. This is due to the availability of a greater variety of (imported) substitutes after opening up to trade, which makes it easier to move away from a good (or reduce its use) on a price increase. According

to Hicks's well-known "fundamental law of factor demand," output-demand elasticities and factor-demand elasticities move in the same direction (see Slaughter 2001 for a more detailed discussion). Thus greater trade openness should lead to an increase in labor-demand elasticities. As Rodrik (1997) notes and as explained in Slaughter (2001) and Krishna, Mitra, and Chinoy (2001), bigger labor-demand elasticities shift the incidence of nonwage labor costs toward workers, lead to wider fluctuations in wages and employment due to fluctuations in labor demand, and reduce the bargaining power of workers. Slaughter (2001) was the first to perform an empirical investigation of this positive relationship between openness and labor-demand elasticities, but only for a developed country, namely the United States. He found mixed empirical support for this relationship using U.S. four-digit industry-level data. The first developing-country study in which this link has been investigated is Krishna, Mitra, and Chinoy (2001). Using a partial equilibrium, imperfectly competitive setup and plant-level data from the Turkish manufacturing sector, they find no linkage between greater trade openness and labor-demand elasticities. However, a recent paper by Hasan, Mitra, and Ramaswamy (2007), using state-level, industry-level data from a large developing country, namely India, finds a statistically significant effect in the predicted direction, the result being stronger in states with more flexible labor markets.

The literature on the link between trade and labor provides evidence that trade affects wages but does not affect industry-level employment due to the presence of various mutually offsetting channels. Trade has also been shown to have increased wage inequality. In addition, a new literature, with mixed evidence on the effects of trade on labor-demand elasticities has emerged.

Finally, there is some recent work on trade openness and child labor, which has not been covered in this entry (see Edmonds and Pavcnik, 2005).

The purpose of this entry is to examine the impact of trade on a few important aspects of development. Growth in incomes is an important component of development. However, if this growth takes place

without any reduction in poverty, an important purpose of development is defeated. Therefore the impact of trade on incomes, growth, and poverty reduction has been discussed in this entry. Additionally, the trade-labor nexus has also been studied, as a large majority in developing countries does not possess any productive assets other than raw labor power.

See also child labor; export promotion; import substitution industrialization; North-South trade; terms of trade; trade and wages

FURTHER READING

Bhagwati, Jagdish N. 1958. "Immiserizing Growth: A Geometrical Note." *Review of Economic Studies* 25: 201 5. Geometrically demonstrates that the negative terms of trade effect (arising from growth) can dominate the di rect effects of growth, leading to a welfare loss.

⸻. 1971. "The Generalized Theory of Distortions and Welfare." In *Trade, Balance of Payments and Growth*, edited by J. N. Bhagwati et al. North Holland, Amsterdam, chapter 12. Looks at the effects of trade policy in the presence of market distortions and im perfections and argues that the positive effects of trade on welfare are restored in the presence of direct correc tive policies.

Broda, Christian, and David E. Weinstein. 2006. "Globa lization and the Gains from Variety." *Quarterly Journal of Economics* 121 (2): 541 85. Uses very disaggregate data to measure the impact of trade on welfare through the variety channel.

Currie, Janet, and Ann E. Harrison. 1997. "Sharing the Costs: The Impact of Trade Reform on Capital and Labor in Morocco." *Journal of Labor Economics* 15 (3): S44 71. Looks at some of the mutually offsetting effects of trade on firm level employment.

Devarajan, Shantayanan, and Dani Rodrik. 1991. "Pro Competitive Effects of Trade Reform: Results from a CGE Model of Cameroon." *European Economic Review* 35 (5): 1157 84. Analyzes the effect of import compe tition on innovation through the flattening of the product demand curve.

Dollar, David, and Aart Kraay. 2002. "Growth Is Good for the Poor." *Journal of Economic Growth* 7 (3): 195 225. Shows empirically that growth reduces poverty.

Edmonds, Eric, and Nina Pavcnik. 2005. "Child Labor in the Global Economy." *Journal of Economic Perspectives* 18 (1): 199 220. Surveys literature on the impact of globalization on child labor.

Feenstra, Robert C., and Gordon H. Hanson. 2003. "Global Production and Inequality: A Survey of Trade and Wages." In *Handbook of International Trade*, edited by E. K. Choi and J. Harrigan. London: Blackwell, 146 85. Surveys how input trade can affect and has affected wage inequality.

Findlay, Ronald E. 1984. "Growth and Development in Trade Models." In *Handbook of lnternational Economics*, vol. 1, edited by R. W. Jones and P. B. Kenen. Amsterdam: North Holland, 185 236. Surveys the trade and growth literature prior to the emergence of endogenous growth theory.

Frankel, Jeffrey, and David Romer. 1999. "Does Trade Cause Growth?" *American Economic Review* 89 (3): 379 99. An empirical, cross country study on trade and growth.

Goldberg, Pinelopi, and Nina Pavcnik. 2007. "Distributional Effects of Globalization in Developing Countries." *Journal of Economic Literature* 45 (1): 39 82. Survey article on the impact of globalization on the various aspects of inequality.

Grossman, Gene, and Elhanan Helpman. 1991. *Innovation and Growth in the Global Economy*. Cambridge, MA: MIT Press. Collection of their theoretical work on trade and endogenous growth.

Harrison, Ann E. 1994. "Productivity, Imperfect Competition, and Trade Reform: Theory and Evidence." *Journal of International Economics* 36 (1 2): 53 73. Empirical plant level study of the impact of trade reform on productivity growth and market structure.

Hasan, Rana, Devashish Mitra, and K. V. Ramaswamy. 2007. "Trade Reforms, Labor Regulations, and Labor Demand Elasticities: Empirical Evidence from India." *The Review of Economics and Statistics* 89 (3): 466 81. Empirical, state level, industry level study of the impact of trade liberalization on labor demand elasticities in the presence of varying labor market institutions.

Hasan, Rana, Devashish Mitra, and Beyza Ural. 2007. "Trade Liberalization, Labor Market Institutions and Poverty Reduction: Evidence from Indian States." *India Policy Forum* 3: 71 122.

Hasan, Rana, M. G. Quibria, and Yangseon Kim. 2003. "Poverty and Economic Freedom: Evidence from Cross Country Data." Economics Study Area Working Papers 60. Honolulu, Hawaii: East West Center. An empirical analysis of how various components of economic freedom affect poverty.

Krishna, Pravin, and Devashish Mitra. 1998. "Trade Liberalization, Market Discipline, and Productivity Growth: New Evidence from India." *Journal of Development Economics* 56 (2): 447 62. Empirical firm level study of the impact of trade reform on productivity growth and market structure, followed by some welfare analysis.

Krishna, Pravin, Devashish Mitra, and Sajjid Chinoy. 2001. "Trade Liberalization and Labor Demand Elasticities: Evidence from Turkey." *Journal of International Economics* 55 (2): 391 409. Empirical plant level study of the impact of trade liberalization on labor demand.

Levinsohn, James. 1993. "Testing the Imports as Market Discipline Hypothesis." *Journal of International Economics* 35 (1 2): 1 22. Empirical plant level study of the impact of trade reform on market structure (the price cost markups).

Pavcnik, Nina. 2002. "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants." *Review of Economic Studies* 69 (1): 245 76. Empirical plant level study of the impact of trade reform on productivity growth, using modern techniques to address various econometric issues.

Porto, Guido G. 2006. "Using Survey Data to Assess the Distributional Effects of Trade Policy." *Journal of International Economics* 70 (1): 140 60. Develops a method to estimate the general equilibrium distributional effects of trade policies using household survey data.

Ravallion, Martin. 2001. "Growth, Inequality and Poverty: Looking Beyond Averages." *World Development* 29 (11): 1803 15. Assesses the degree to which poor people share in growth.

Rodriguez, Francisco, and Dani Rodrik. 2001. "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross National Evidence." In *Macroeconomics Annual 2000*, edited by Ben Bernanke and Kenneth S. Rogoff. Cambridge, MA: MIT Press for NBER, 261 325. Economic and econometric critique of cross country

empirical work showing the positive effects of trade on growth.

Rodrik, Dani. 1992. "Closing the Productivity Gap: Does Trade Liberalization Really Help?" In *Trade Policy, Industrialization, and Development: New Perspectives*, edited by G. K. Helleiner. Oxford: Clarendon Press, 155 75. Analyzes how import competition affects growth by reducing the market size of domestic, import competing firms.

———. 1997. *Has Globalization Gone Too Far?* Washington, DC: Institute for International Economics. A detailed analysis of the costs and benefits of international economic integration.

Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi. 2004. "Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development." *Journal of Economic Growth* 9 (2): 131 65. Empirically investigates the interactions between trade, institutions, and geography and their effects on incomes.

Romer, Paul M. 1986. "Increasing Returns and Long Run Growth." *Journal of Political Economy* 94 (5): 1002 37. Pathbreaking paper introducing and analyzing the concept of endogenous growth.

Sachs, Jeffrey D., and Andrew Warner. 1995. "Economic Reform and the Process of Global Integration." *Brookings Papers on Economic Activity*: 1 118. Studies the impact of economic reform on growth.

Slaughter, Matthew. 2001. "International Trade and Labor Demand Elasticities." *Journal of International Economics* 54 (1): 27 56. Empirical investigation of the effects of trade on labor demand elasticities using U.S. industry level data.

Topalova, Petia. 2005. "Trade Liberalization, Poverty, and Inequality: Evidence from Indian Districts." In *Globalization and Poverty*, edited by Ann Harrison. Cambridge, MA: University of Chicago Press and National Bureau of Economic Research. Empirical, district level study of the impact of trade liberalization on poverty in India.

Winters, L. Alan, Neil McCulloch, and Andrew McKay. 2004. "Trade Liberalization and Poverty: The Evidence So Far." *Journal of Economic Literature* 42 (1): 72 115. Surveys the empirical literature on trade and poverty.

DEVASHISH MITRA

trade and the environment

Trade, the exchange of goods and services across countries, is often viewed as an engine of economic growth. Benefits of liberalized trade include access to a larger variety of goods and services to consumers, easier access to foreign technologies, access to larger markets for producers, and increased efficiency in resource allocation. The impact of trade on the environment, however, is a contentious issue; air and water pollution, the degradation of natural habitats, loss of species, and global pollutants, particularly carbon dioxide emissions, are major concerns.

Recent Trends In the late 20th and early 21st centuries international trade has rapidly increased worldwide, while average tariffs and quantitative restrictions to trade (import and export quotas) have fallen steadily. Export growth has outpaced growth in gross domestic product (GDP) (World Bank 2006). With respect to the environment, there are persistent and widespread improvements in most *local* urban pollutants, mainly airborne pollutants (e.g., sulfur dioxide, nitrogen dioxide, carbon monoxide, air particles, and lead). An important exception is local ozone, a highly dangerous local pollutant that has increased over time in most cities, in part as a consequence of measures taken to reduce some other air pollutants. There are also less clear but perceptible trends to improve some indicators of water quality (United Nations 2006). In contrast, there is a clear worsening of the *global* pollutants (e.g., carbon dioxide) as well as of the rural or "green" environment (i.e., the natural forests and other important natural habitats). The latter phenomenon is causing a precipitous loss of species and has contributed to increased global warming through the emissions of carbon dioxide due to massive forest burning.

Thus while trade has rapidly expanded, the environmental trends show a sharp dichotomy: the local urban environment improves but the green or rural environments continuously deteriorate. One reason for this may be that local urban pollution is directly felt by large concentrations of population able to exert strong pressures on politicians to control it, while the rural environment affects directly a smaller fraction of the population which, due to its

geographical dispersion, is less able to pressure governments. Rural environmental degradation and global pollution are less visible to the majority of the population than local urban pollution, which might explain the generally more lax response by governments to rural than urban environmental problems and to global rather than local pollutants.

Effects of Trade on the Environment The central issue is whether opening up to trade has magnified the trends described above or has instead mitigated some of them. The effects of trade on the environment can be broken down into scale, technique, composition, and growth effects (Antweiler, Copeland, and Taylor 2001; López, Galinato, and Islam 2007).

Scale effect. Most forms of pollution are a by-product of a production process. Increased trade openness often implies an increase in economic activity. The scale effect, holding constant production techniques and the mix of goods produced, is likely to cause an increase in the level of local and global pollution and also faster degradation of natural resources. For example, expanding agricultural exports may increase agricultural activities, which may result in water pollution from extensive fertilizer use and deforestation from increased demand for agriculture. The scale effect may also include trade-related direct increases in pollution emissions through increase in air and road transportation. Empirical studies usually employ gross domestic product (GDP) per square kilometer as a proxy for the scale effect. López (1997), for example, found that in Ghana trade liberalization induced a faster rate of deforestation. Given the more lenient attitude of governments toward rural environmental degradation and global pollutants than to urban local pollution, the negative impact of the scale effect is likely to be worse for the green-global environment than for the urban environment.

Technique (wealth) effect. The technique effect refers to reductions in emission intensity per unit of output. If trade raises income, emission intensity may fall if environmental quality is a normal good. A normal good is one for which, as incomes rise, individuals would prefer more of it. Higher income may lead to stricter environmental regulation, under

the assumption that country governments are responsive to the citizens' demands. A trade-induced rise in incomes would thus make higher environmental quality desirable. Empirical studies often use per capita GDP as a proxy for income. A more accurate measure is per capita household consumption expenditure, which is more directly related to permanent income or wealth than per capita GDP (López, Galinato, and Islam 2007). The technique effect of trade has been found to reduce certain pollutants, particularly air pollutants, but the effects on other environmental factors is less significant. The strength of the technique effect is weaker for the green and global environment than for the local urban environment because the citizens' demands for environmental quality are likely to be feebler in the rural areas and for global pollutants than for the control of local urban pollutants.

Composition effect. Trade may also alter the composition of the economy's output. If the economy's comparative advantages favor clean industries, increasing trade openness may switch from pollution-intensive "dirty" goods to less polluting, or "clean," goods and services. The general assumption is that production of dirty goods is more intensive in physical capital and natural resources while clean goods production is more intensive in human capital. Holding the scale of production and other factors constant, an economy that shifts its production toward physical capital intensive goods will pollute more, and conversely, an economy that shifts its production away from physical capital intensive goods will pollute less. Countries that have large endowments of natural resources are likely to relatively specialize in resource-intensive industries and thus increase the extraction of natural resources when they open to trade. In countries where property rights on resources are poorly defined or where environmental regulations are not properly enforced, increased trade is likely to result in more resource degradation and deforestation.

Even more seriously, lack of property rights on resources may lead countries to specialize in natural resource intensive activities and hence to further environmental degradation even if they are not richly

endowed in resources. That is, the institutional and regulatory failures may lead to false comparative advantages, in which case trade may reduce rather than raise income as is normally assumed (this is behind the *pollution haven hypothesis* as discussed later). In this perverse case the technique effect discussed earlier (which assumed that income increases with trade) would be reversed. Also, once again, the weight of the composition effect may be felt on the green-global environment because environmental control institutions and regulatory policies are less developed for the rural-global resources than for urban pollutants.

Growth rate effect. Trade openness may cause a number of dynamic forces that promote not only a once-and-for-all effect on the income level but also a faster pace of economic growth over time. For example, trade openness may cause an economy to adopt new technologies at a faster rate due to the fact that many new technologies are generated abroad. A faster pace of economic growth may cause lower environmental quality than a country growing at a slower rate (López, Galinato, and Islam 2007). The issue here is that environmental institutions and policies need time to be adapted. An economy growing at a fast rate will find it much more difficult to timely adapt their policies and institutions to properly respond to increasing pollution than an economy growing at a more moderate pace. This trade-induced growth rate effect may result in a decline in environmental quality.

The net effect. Empirical studies seem to corroborate the hypothesis that the positive-technique effect dominates the other effects for certain local urban pollutants resulting in trade being good for the urban environment (Copeland and Taylor 2003). However, the few empirical studies of the impact of trade on the rural environment, particularly the impact on wetlands and natural forests, suggest that the net effect of trade is negative (López 1997). This is consistent with the earlier conceptual discussion regarding the relative strength of the various partial effects on the urban and rural environments. The net effect of trade is also likely to be negative for global pollutants because the technique effect may be weak

for such pollutants as people care less for pollutants that do not affect them directly.

Pollution Haven Hypothesis According to this hypothesis the direction of trade between two countries may be dominated by differences in environmental regulatory strengths. Developed economies tend to have better environmental institutions and more efficient regulation than poorer countries. The pollution haven hypothesis states that rich countries may export their dirty industries to poorer countries due to the differences in regulation. Freer trade thus results in declining environmental quality in poorer economies and improving environmental quality in richer ones (Chichilnisky 1994). There are two main assumptions behind the pollution haven hypothesis: first, that pollution regulation differences are a key determinant of industry location, and second, that environment is a normal good and thus differences in regulation are due to income differences. A further implication of this hypothesis is that global environmental quality may deteriorate and the income of the poorer economies may fall. As polluting industries migrate to regions with less stringent pollution policy, overall global pollution will increase. The available empirical evidence generally rejects the pollution haven hypothesis, but this does not mean that environmental regulation plays no role in affecting trade (Copeland and Gulati 2006). The evidence simply says that environmental regulatory differences do not necessarily dominate the direction of trade as the pollution haven hypothesis suggests.

Factor Endowment Hypothesis The factor endowment hypothesis deviates from the pollution haven hypothesis by postulating that factor endowments, and not just differences in environmental regulations, are the main motivation for trade patterns. Economies engaging in trade will specialize in production where comparative advantage is exhibited. If the most developed countries are relatively abundant in factors (usually capital) used in pollution-intensive industries, then they may have a comparative advantage in dirty industries and thus will specialize in them. Consequently, dirty goods production may shift from developing to developed economies. Developed economies have better

environmental regulation and institutions, and thus the consequence of trade would be an overall decline in pollution. Proponents of this hypothesis point to the fact that Europe and the United States have the most stringent pollution policies yet export manufactured goods that are highly pollution intensive. If developing countries have an abundance of the factor needed by pollution-intensive production, the predictions of the factor endowment hypothesis would be consistent with the pollution haven hypothesis. Furthermore, developed countries will lose their comparative advantage in dirty industries if stringency in pollution policy is increased. Thus there are two forces at play, pollution policy differences and factor endowments.

Trade Openness and the Environmental Kuznets Curve Part of the empirical environmental literature has been on the relationship between income and pollution, otherwise known as the Environmental Kuznets Curve (EKC) (Shafik and Bandyopadhyay 1992; Grossman and Krueger 1995). The EKC is an inverted U-shaped relationship: as income increases, pollution first increases until it reaches a turning point and then declines. The first empirical estimation of the EKC for air and water pollutants was carried out by Grossman and Krueger (1995). Estimation of the EKC has also been carried out for natural resources such as forests (López and Galinato 2005). There is much debate with regards to the empirical estimates, data accuracy, robustness, and theoretical underpinnings of the EKC (Harbaugh, Levinson, and Wilson 2000; Deacon and Norman 2007).

Trade plays an important role in some of the conceptual explanations of the EKC. The income effect theory identifies environmental policy response as the main reason for the EKC. As an economy grows, at first the benefits of increasing output are so large that they dominate the increased demands for environmental quality caused by a higher income and thus the scale effect dominates (the curve is upward sloping in this segment). Beyond a certain level of income, the marginal preference for more consumption declines and the preference for clean environment increases until the turning point occurs. Pollution declines as income

increases beyond this point. What trade does is to enhance the process of economic growth and thus has an indirect effect on the EKC.

Another story focuses on the composition effect. In the early stages, countries grow through physical capital accumulation and expansion of industries intensive in physical capital, which are generally dirty; in the latter stages, a country grows through human capital and knowledge accumulation and thus cleaner industries emerge, yielding an EKC relationship. Trade liberalization may assist in the switch from dirty to clean industries by allowing the growing economy to increasingly specialize in clean industries. Without trade the assumptions required for an EKC process are much more stringent than with trade.

The Resource Curse Countries with larger endowments of natural resources apparently tend to grow less than resource-poor countries (Sachs and Warner 1999). There are several explanations for such a phenomenon (Barbier 2005). One of the most credible explanations is directly linked with trade, the so-called Dutch disease effect. As a natural resource dependent economy booms, resources are allocated from other sectors to the natural resource sector. Furthermore, the currency of the economy appreciates, which renders other sectors in the economy uncompetitive in international markets. This results in further dependency on the natural resource sector. Hence the economy is more vulnerable to the price fluctuations inherent to primary commodities. Examples of the resource curse can be seen in oil-producing countries that are resource rich but are growing slowly.

The Role of Government Environmental quality is a public good; thus by definition the market will underprovide it. Although much emphasis has been placed on government efficiency and the provision of public goods such as environmental quality, there has been less emphasis placed on the efficiency of government subsidies. Governments that provide trade and other subsidies do so at the expense of underproviding public goods, given a fixed budget constraint. Public good investments by the government can complement private investments and alleviate

market failures, thus resulting in economic growth (López and Miller 2007; López and Galinato 2007). Furthermore, government subsidies, including trade subsidies, would promote activities that would be more demanding for the environment as opposed to public good expenditures, which may compensate for credit market failures and promote human capital accumulation. A study by López, Galinato, and Islam (2007) finds that increasing the share of public goods in total government expenditures reduces SO_2, NO_2, and lead pollution.

Empirical and conceptual analyses suggest that trade has contributed to economic growth and has accelerated trends to ameliorate local air pollutants. This is particularly true of most local air pollutants affecting cities. Trade does not seem to mitigate the ever-increasing emission of global pollutants, particularly carbon dioxide. The few studies focused on the links between trade and the green environment suggest that increased trade appears to exacerbate the losses of natural forests and other natural habitats, thus aggravating the trends toward global climate change and loss of biodiversity.

See also Basel Convention; Convention on Biological Diversity; Convention on International Trade in Endangered Species (CITES); Global Environment Facility; multilateral environmental agreements; pollution haven hypothesis

FURTHER READING

Andreoni, James, and Arik Levinson. 2001. "The Simple Analytics of the Environmental Kuznets Curve." *Journal of Public Economics* 80 (2): 269 86. A theoretical and empirical analysis of the role of economies of scale of abatement technology on the Environmental Kuznets Curve.

Antweiler, Werner, Brian R. Copeland., and M. Scott Taylor. 2001. "Is Free Trade Good for the Environment?" *American Economic Review* 91 (4): 877 908. A rigorous theoretical and empirical examination of the relationship between trade and the environment.

Barbier, Edward B. 2005. *Natural Resources and Economic Development*. Cambridge: Cambridge University Press. A thorough review and exploration of the relationship between natural resources and development.

Chichilnisky, Graciela. 1994. "North South Trade and the Global Environment." *American Economic Review* 84 (4): 851 74. An important theoretical paper on the relationship between trade and property rights.

Copeland, Brian, and Sumeet Gulati. 2006. "Trade and the Environment in Developing Countries." In *Economic Development and Environmental Sustainability*, edited by R. López and M. Toman. Oxford: Oxford University Press, 178 216. An exploration of the relationship among capital mobility, international trade, and environmental quality from a developing country perspective.

Copeland, Brian R., and M. Scott Taylor. 2003. *Trade and the Environment*. Princeton, NJ: Princeton University Press. A comprehensive theoretical and empirical exploration of several aspects of the relationship between trade and the environment.

Deacon, Robert T., and Catherine Norman. 2007. "Is the Environmental Kuznets Curve an Empirical Regularity?" In *Frontiers of Environmental and Natural Resource Economics*, edited by R. Halvorsen and D. Layton. UK: Edward Elgar. An empirical analysis of the robustness of the Environmental Kuznets Curve.

Grossman, Gene M., and Alan B. Krueger. 1995. "Economic Growth and the Environment." *Quarterly Journal of Economics* 112 (2): 353 78. One of the earliest examinations of the Environmental Kuznets Curve for a wide range of environmental indicators.

Harbaugh, William, Arik Levinson, and David Wilson. 2000. "Reexamining the Empirical Evidence for an Environmental Kuznets Curve." *Review of Economics and Statistics* 84 (3): 541 51. An empirical analysis update of the Environmental Kuznets Curve using revised data subject to various specification alterations.

López, Ramón. 1997. "Environmental Externalities in Traditional Agriculture and the Impact of Trade Liberalization: The Case of Ghana." *Journal of Development Economics* 53 (1): 17 39. A theoretical and empirical examination of the impact of trade policies on environmental resources in Ghana.

Lopez, Ramón. and Gregmar I. Galinato. 2005. "Deforestation and Forest Induced Carbon Dioxide Emissions in Tropical Countries: How Do Governance and Trade Openness Affect the Forest Income Relationship?" *Journal of Environment and Development* 14 (1): 73 100. Studies the implications of changes in land use induced

by economic growth, economy wide policies, and gov
ernance on deforestation and forest induced atmo
spheric carbon dioxide emissions.

. 2007. "Should Governments Stop Subsidies to
Private Goods? Evidence from Rural Latin America."
Journal of Public Economics 91 (5 6): 1071 94. An
important analysis of the impact of government sub
sidies on per capita income, poverty, and environment in
rural Latin America.

López, Ramón, Gregmar Galinato, and Asif Islam. 2007.
"Government Expenditures and Air Pollution." Uni
versity of Maryland at College Park Working Paper. An
empirical analysis of the relationship between the share
of government expenditures in public goods and air
pollution.

López, Ramón, and Sebastian Miller. 2007. "The Structure
of Public Expenditure: A Robust Predictor of Economic
Development?" University of Maryland at College Park
Working Paper. A thorough cross country analysis of the
impact of the share of government expenditures in
public goods on economic growth.

Sachs, Jeffrey D., and Andrew W. Warner. 1999. "The Big
Push, Natural Resource Booms and Growth." *Journal of
Development Economics* 59 (1): 43 76. A theoretical and
empirical examination of the relationship between nat
ural resource booms and economic growth for Latin
American countries.

Shafik, Nemat, and Sushenjit Bandyopadhyay. 1992.
"Economic Growth and Environmental Quality: Time
Series and Cross Country Evidence." Policy Research
Working Paper No. WPS 904. Washington, DC:
World Bank. One of the earliest studies on the re
lationship between economic growth and environ
mental quality, later to be dubbed the Environmental
Kuznets Curve.

United Nations. 2006. *Trends in Sustainable Development.*
New York: UN Department of Economic and Social
Affairs. An overview of global trends in sustainable de
velopment indicators.

World Bank. 2006. *Assessing World Bank Support for Trade,
1987 2004: An IEG Evaluation.* Washington, DC:
World Bank. An evaluation of the World Bank's in
volvement and support for international trade.

RAMÓN LÓPEZ AND ASIF ISLAM

■ trade and wages

Beginning with Adam Smith and David Ricardo, economists have demonstrated the benefits of free trade in numerous ways, yet free trade has hardly become a universally accepted strategy. Even in cases where free trade maximizes a country's welfare (or generates a higher welfare than trade restriction), it is not necessarily a superior policy. Free trade creates both winners and losers in a trading nation. It is only through a compensation principle (by which the winners are forced to compensate the losers) that everyone can be made better off under free trade. In practice, however, such a compensation principle is seldom applied; more relevant, then, is how trade affects factor incomes or wages and thus the income distribution within a country. The study of trade and wages is important for two other reasons. First, as most of the poor belong to the working class, study of the effects of trade on wages helps us understand how trade alleviates poverty in the short term. Second, if free trade widens income inequality, it fuels interclass conflict. This makes it difficult for a democratic government to pursue free trade as a long-term de-velopment policy.

Theory The best-known result regarding the static effect of trade liberalization on wages is the Stolper-Samuelson (SS) theorem, or its more general version known as the price magnification effect (Jones 1965): In a two-commodity two-factor (2x2) world, a decline in the (domestic) price of imports relative to that of the exports consequent on, say, trade liberalization leads to a more than propor-tionate fall in the rate of return to the factor used intensively in the import-competing sectors and a more than proportionate increase in the rate of return to the other factor used intensively in the export sectors. Thus one factor of production is unambig-uously better off and the other unambiguously worse off in terms of their *real* returns.

Factor-intensity ranking of goods does not mat-ter, however, in a two-commodity three-factor setup of the specific-factor model, or in a 2x2 Heckscher-Ohlin (HO) model with sectorally immobile capital. The factor specific in the export sector gains unam-biguously whereas the factor specific in the import-

competing sector loses unambiguously in both absolute and real terms when trade is liberalized. The mobile factor, on the other hand, gains only when it spends a greater proportion of its income on the import-competing good than on the export good.

The impact of trade liberalization on income distribution is less clear in a many-good many-factor scenario. If goods are ranked and indexed according to the descending order of magnitude of their price changes, the price magnification effect at most tells us that there exists one factor h that is unambiguously better off and one factor k that is unambiguously worse off (Samuelson 1953). But not much can be said about changes in the rate of return to other factors of production in particular, whether they are rising or falling with trade liberalization. The natural enemy and natural friend propositions of Jones and Scheinkman (1977), however, provide us some direction in this regard. Commodity j is said to be a natural enemy (friend) of factor h if an increase in its price, all else being equal, lowers (raises more than proportionately) the rate of return to factor h. But whereas each factor has at least one natural enemy, there may not be any natural friend. The strongest resemblance of the 2x2 SS theorem can be found in Jones and Marjit (1985), which identifies for each factor a good as a natural friend and all others as natural enemies. It provides a natural sufficiency condition for the theorem in higher dimensions.

All these variants of the SS, or price magnification, effects, however, predict that opening up or multilateral liberalization of trade among nations leads to asymmetric wage movements. The wage-rental ratio should rise in the country exporting relatively labor-intensive commodities and should decline in its trading partner that exports relatively capital-intensive commodities. In a reinterpretation of capital as human capital, trade liberalization, therefore, should widen the wage gap between skilled and unskilled workers in one country and reduce it in the other. These asymmetric wage movements are due to the fact that starting from autarky the domestic prices move asymmetrically in trading nations to converge toward an intermediate equilibrium once trade opens up.

There are, however, two special cases where wages move in the same direction. First is the case of factor-intensity reversal with large difference in the factor-endowment ratios of the trading nations. Thus, for example, though the price of good X relative to that of good Y rises in one country and falls in another, since given the endowment patterns the same good X has different intensity ranking in the two countries, by the price magnification effect the wage-rental ratio (or the wage gap in a human capital reinterpretation) should move in the same direction in both the countries.

Second is the case of Metzler's paradox where deterioration of the terms of trade is so large after a tariff reduction by a home country as to raise, rather than lower, the tariff-inclusive domestic price, which can occur when the price elasticity of foreign import demand is smaller than home consumers' marginal propensity to consume the export good. With the foreign country keeping its (nonprohibitive) tariff rate unchanged, the domestic prices rise in both the countries and thereby produce symmetric wage movements.

Recent developments in theory, however, have provided us with three more *general* cases in which trade and investment liberalization can produce *symmetric* wage movements across the trading nations. In particular, as a response to the observed fact that relative wages of the unskilled have worsened all across the globe, such models bring in the trade pattern and structural characteristics of the developing world.

The first is the case of the generalized HO model with three goods but only two factors of production discussed in Marjit and Acharyya (2003) and Davis (1996). With different cones of diversification for North and South due to asymmetric endowment patterns, the middle good Y is ranked differently in two countries. Therefore, one can get symmetric movement in wages as effectively factor intensities differ. A large capital inflow from the capital-rich country to another (the host) country has the same effect, because such flows induce the host to specialize completely in the most capital-intensive good and relocate the global production of good Y entirely in the host country, thereby raising the demand for

capital (relative to labor) everywhere. In a human capital reinterpretation, therefore, the skilled-unskilled wage gap should increase in both countries in such a generalized HO setting. At the same time the framework can be used to explain asymmetric wage movements following trade liberalization among the Southern countries (Davis 1996). Whereas symmetric changes in wage inequality in both North and South follow from the *local factor intensity ranking*, the asymmetric effect across the Southern countries follows from *local factor abundance*.

The other two cases arise in the context of trade in intermediate products. Using a model with a continuum of stages of production, Feenstra and Hanson (1996) provide an interesting case in the context of the North American Free Trade Agreement and foreign investment. An inflow of U.S. capital into Mexico transfers some production activities from the United States to Mexico that are more skilled-labor-intensive in Mexico but less skilled-labor-intensive in the United States. Therefore relative demand for skilled labor increases in both countries, resulting in growing wage inequality across the border. Marjit and Acharyya (2006), on the other hand, present a more traditional production structure, with one homogeneous intermediate good being produced by skilled labor, which is then combined with skilled labor to produce the manufacturing good. This possibly captures the case of the computer industry. Some of the Southern countries such as India have emerged as phenomenal exporters of software-related products. But software sectors depend on hardware capabilities. If computers are imported from the North, a decline in the import tariff in the South must help the skilled labor in the software-processing industries in the South. But a lower tariff on hardware implies that items that are skilled-labor-intensive will be more profitable to produce in the North. This tends to increase wages for skilled labor across the globe. Marjit, Beladi, and Chakrabarti (2003), Jones and Marjit (2003), and Marjit and Kar (2005) present cases explaining how declining information costs, capital inflow, and unskilled labor outflow can lead to a rise in the relative wage of skilled labor in a developing economy.

Evidence: Developed Countries Since the late 1970s there has been a dramatic increase in the inequality of earnings based on education, experience, and occupation in the United States. The ratio of the average wage of a college graduate to that of a high school graduate rose by 15 percent. Leamer (2000) also observed a steady and persistent widening, since 1971, of the earnings gap in the manufacturing sector between college graduates and high-school dropouts. The 1994 share of salary of high-school dropouts compared with college graduates was only 15 percent of its 1971 level. For high-school graduates the situation improved during the first half of the 1970s, but after that deteriorated steadily. Similar increases in wage inequality during the 1980s and 1990s have been observed for Australia, Canada, Japan, and Europe. Between 1978 and 1988, the ratio of manual-labor to nonmanual-labor wages fell by 8.1 percent in Germany and by 3 percent in Italy whereas it actually increased in Belgium and Denmark. In the United Kingdom there was a substantial increase in the ratio of earnings of the highest to lowest percentiles.

Such a dramatic change in relative wages has triggered heated discussions and debates on the causes of these changes (Krugman 2000; Jones and Engerman 1996; Leamer 2000). Essentially, two sets of issues have become dominant in this debate: trade-related and technology-related. Though this debate is yet to be settled once and for all, to many researchers global competition has bifurcated American workers into two groups: the high-earning "symbolic analysts," whose talents are rewarded by globalization, and the mass of ordinary production workers, whose earnings are depressed by it. Similarly in the case of Europe, downward competitive pressures on working-condition standards as a result of freer trade with Eastern Europe and Asia have been conceived as the major cause of low wages. Moreover the accession of low-wage countries such as Spain, Portugal, and Greece into the European Union has worsened the scenario through relocation of firms to these countries. The relocation of the Hoover Corporation from France to Scotland in the year 1988, attracted by both lower wage costs and lower working-condition standards, is a glaring example. More recently, com-

paring merchandise trade with merchandise value added, Feenstra and Hanson (1996) have observed that, except for Australia, Japan, and the United Kingdom, all other industrial countries experienced substantial growth in trade in 1990 than in 1913.

Developing Country Experience There are quite a few systematic studies on the impact of liberal trade policies on internal distribution of wage and income in the developing world. Robbins (1995) demonstrated that while wage inequality decreased a bit in the East Asian nations, Latin America in general experienced a widening wage gap between the skilled and the unskilled after trade liberalization. Wood (1997) argued why such evidence was untenable in terms of traditional trade theoretic arguments. We have mentioned how subsequent theoretical research has used variants of standard trade models to justify double-sided wage gap outcome. A more detailed survey of the related empirical literature is available in Marjit and Acharyya (2003). Whether it is trade or technology that lies at the root of rising income inequality in the developing world is still very much an open question.

Trade and Skill Formation Globalization has meant a remarkable degree of economic integration, thanks to declining tariff walls and revolutionizing changes in information technology. There is a general tendency for the process to reward skills that tend to flourish through international trade and investment. Hence skill formation becomes the most fundamental driving force determining local as well as global patterns of income distribution. An enduring line of research will be whether trade promotes or inhibits acquisition of skills that are adaptable to changes in the global environment. Early attempts to model these are made in Marjit and Acharyya (2003) and Acemoglu (2003).

See also Heckscher-Ohlin model; International Labor Organization; labor standards; specific-factors model

FURTHER READING

Acemoglu, Daron. 2003. "Patterns of Skill Premia." *Review of Economic Studies* 70 (2): 231 51. Discusses skill acquisition and wage inequality in North and South.

Davis, Donald. 1996. "Trade Liberalization and Income Distribution." NBER Working Paper No. 5693. Cambridge, MA: National Bureau of Economic Research. A lucid theoretical illustration of asymmetric changes in the wage gap in the developing world.

Feenstra, Robert, and Gordon Hanson. 1996. "Foreign Investment, Outsourcing, and Relative Wages." In *Political Economy of Trade Policies: Essays in Honour of J. N. Bhagwati*, edited by Robert Feenstra, Gene Grossman, and Douglas Irwin. Cambridge, MA: MIT Press, 89 127. Establishes the two way widening wage gap due to foreign investment in a Ricardian model with continuum of stages of production.

Jones, Ronald Winthrap. 1965. "The Structure of Simple General Equilibrium Models." *Journal of Political Economy* 73: 557 72. This pioneering and most cited work provides the basic analytical structure and algebraic exposition of 2x2 general equilibrium trade models and their properties.

Jones, Ronald Winthrap, and Stanley Engerman. 1996. "Trade, Technology, and Wages: A Tale of Two Countries." *American Economic Review* 86: 35 40. Emphasizes the one to one correspondence between commodity and factor prices in an open economy that is often misunderstood by labor economists and summarizes the relevance of the technology and trade explanations in the wage gap debate.

Jones, Ronald Winthrap, and Sugata Marjit. 1985. "A Simple Production Model with Stolper Samuelson Property." *International Economic Review* 26 (3): 565 67. Provides a sufficient condition and a related simple production structure for higher dimensional proof of the Stolper Samuelson theorem.

———. 2003. "Economic Development, Trade, and Wages." *German Economic Review* 4 (1): 1 17. Brings agriculture into the wage inequality literature and shows that a price rise in agriculture may still widen the wage gap between the skilled and unskilled.

Jones, Ronald Winthrap, and Jose Scheinkman. 1977. "The Relevance of the Two Sector Production Model in Trade Theory." *Journal of Political Economy* 85: 909 35. Extends the price magnification effect to n commodity n factor cases and derives the natural enemy and natural friend propositions.

Krugman, Paul. 2000. "Technology, Trade, and Factor Prices." *Journal of International Economics* 50 (1): 51 71. Illustrates how one can isolate the extent of world price

changes (and hence the extent of the rise in wage gap) that are caused by international trade per se in terms of the "but for" question.

Leamer, Edward. 2000. "What's the Use of Factor Contents?" *Journal of International Economics* 50: 51 71. Brings out the limitation of the factor content approach to trade in isolating the impact of interna tional trade from that of technological change on the rising wage gap.

Marjit, Sugata, and Rajat Acharyya. 2003. *International Trade, Wage Inequality, and the Developing Economy: A General Equilibrium Approach*. Heidelberg: Physica/ Springer Verlag. Provides a brief documentation of the wage gap phenomenon, a brief survey on the trade versus technology debate, as well as analytical structures that can explain the rising wage gap in the North and the South alike caused by international trade.

——. 2006. "Trade Liberalization, Skill linked Inter mediate Production, and Two Sided Wage Gap." *Journal of Policy Reform* 9 (3): 203 17. Provides an al ternative theoretical explanation of the two way wage gap caused by trade liberalization in a general equilib rium framework with trade in intermediate good.

Marjit, Sugata, Hamid Beladi, and Avik Chakrabarti. 2003. "Trade and Wage Inequality in Developing Countries." *Economic Inquiry* 42 (92): 295 303. Discusses the im pact of fragmentation on wage inequality.

Marjit, Sugata, and Saibal Kar. 2005. "Emigration and Wage Inequality." *Economics Letters* 88 (1): 141 45. Shows that emigration of unskilled labor does not nec essarily increase relative wage of the unskilled.

Robbins, Donald. 1995. "Trade, Trade Liberalization, and Inequality in Latin America and East Asia: Synthesis of Seven Country Studies." Mimeo, Harvard Institute of International Development. Studies the Latin American and East Asian experiences.

Samuelson, Paul. 1953. "Prices of Factors and Goods in General Equilibrium." *Review of Economic Studies* 21: 1 20. The foremost analysis that extended the Stolper Samuelson and factor price equalization results to many commodity many factor cases.

Wood, Adrian. 1997. "Openness and Wage Inequality in Developing Countries: The Latin American Challenges to East Asian Conventional Wisdom." *World Bank Re search Observer* 11 (1): 33 57. Argues why the Latin

American experience is untenable in terms of the tradi tional trade theoretic arguments.

Xu, Bin. 2003. "Trade Liberalization, Wage Inequality, and Endogenously Determined Nontraded Goods." *Journal of International Economics* 60 (2): 417 31. Explores the implication of nontraded production and the role of trade liberalization in widening of the wage gap.

SUGATA MARJIT
AND RAJAT ACHARYYA

■ trade costs and foreign direct investment

Foreign direct investment (FDI) takes place when a firm from one country sets up a plant or acquires a plant in another country. This entry explores the relationship between trade costs and various types of FDI, such as horizontal FDI, export-platform FDI, mergers and acquisitions, and vertical FDI. Trade costs refer to all costs incurred in getting a good from its production location to the final destination. They include transportation costs (both freight costs and time costs), policy barriers (tariffs and nontariff barriers), information costs, contract enforcement costs, costs associated with the use of different cur rencies, legal and regulatory costs, and local distri bution costs (wholesale and retail). Mirroring the literature, this entry focuses on tariffs and transpor tation costs but also discusses the impact on FDI of information costs, contract enforcement costs, and costs associated with the use of different currencies.

Horizontal FDI Horizontal FDI refers to the case where a multinational firm produces and sells its product in a foreign market. The standard frame work to study horizontal FDI is the ownership-location-internalization (OLI) framework of Dun ning (1973). It suggests that a multinational firm wanting to sell goods in a foreign market faces a choice between exporting and FDI. The key trade-off involved in this choice is between the economies of scale in production and the trading costs of shipping goods to a foreign market, known as the proximity-concentration trade-off. While economies of scale in production favor exporting to the foreign market from the home production facility, a high trading

cost of exporting to the foreign market favors FDI. The implication is that, for a given level of economies of scale, the higher the trading cost the greater the likelihood of FDI occurring. FDI in this framework acts as a substitute for trade, that is, increased FDI involves decreased exports. Brainard (1997) provides empirical support for this finding using data on U.S. multinationals: local production by the affiliates of U.S. multinationals relative to exports from the U.S. parent increases the higher the transport costs and trade barriers and the lower the plant level economies of scale.

Licensing versus Horizontal FDI While the proximity-concentration trade-off framework focuses on the choice between exports and FDI, in the OLI framework, once a multinational firm decides to produce the good locally rather than export it from its home production facility, it faces a choice between licensing the technology to a local firm and undertaking FDI. Since licensing involves arm's length transaction between two independent parties, contracting issues become important. The licensor and the licensee have different sets of information, which gives rise to the standard asymmetric information problem. For example, the licensee has a better idea about the local market conditions, while the licensor knows more about the technology. Neither party has an incentive to reveal its private information to the other party, and the acquisition of more information is costly for each party. In addition, the writing of a contract and its enforcement are costly as well. Therefore, the choice between licensing and FDI depends on these contracting costs. If these costs are high then there are internalization advantages, that is, a firm prefers to keep production internal by undertaking FDI rather than licensing its technology to a local firm. Therefore, while trade costs in the form of tariffs and transportation costs affect the location choice of production, costs arising due to informational asymmetry and contract enforcement determine the choice between production within the firm boundary (FDI) and licensing. All other things being equal, a reduction in the costs of information and contract enforcement would lead to more licensing and less FDI.

The proximity-concentration trade-off framework implies that a decrease in tariffs and transportation cost would reduce FDI, but this prediction is not borne out by the recent experience of European Union (EU) countries, where trade liberalization has been accompanied by increased FDI. One must consider other types of horizontal FDI (and other types of FDI) that could increase with trade liberalization.

Export-Platform FDI If two countries A and B preferentially liberalize trade with each other while keeping the trade barriers on the rest of the world intact, then the firms in a third country C will have an incentive to set up a plant in either country A or B and sell in both those markets from its single production facility. In this case, FDI by firms from country C into country A or B is not just meant for local sales but also serves as its export platform for sales to the other country. Ranjan (2006) shows that preferential trading agreements can create new investment as well as divert investment from nonmembers to members, thereby increasing FDI in the member countries. Ekholm, Forslid, and Markusen (2007) show that a free trade agreement between a high-cost and a low-cost country generates an incentive to an outside high-cost country to set up a plant in the low-cost country within the free trade area. Therefore, mutual trade liberalization within a group of countries can attract FDI into the group from outside countries.

Mergers and Acquisitions A multinational firm building a new plant in a foreign country is called a greenfield FDI, while if instead it acquires an existing plant, then it is classified under mergers and acquisitions (M&A). There is evidence from UNCTAD (2000) that a large amount of FDI is in the form of M&A rather than greenfield. M&As take place for both efficiency and strategic reasons. Efficiency gains come from the fact that a multinational firm may be able to produce at a lower cost by acquiring an existing firm rather than setting up a new plant. Strategic gains from acquiring an existing firm stem from the possibility that profits may increase due to a reduction in the number of firms in the market.

Focusing solely on strategic issues, Neary (2007) shows that a decrease in trade costs may increase the

incentive for M&A. To see the intuition behind this result, suppose there are two countries, Home and Foreign, with oligopolistic market structures. In the case of free trade, only low-cost firms in each country survive. In the presence of high trade costs, relatively higher-cost domestic firms also survive because foreign firms have to bear the trade cost in addition to their production cost. A move from autarky to free trade would make many high-cost domestic firms in each country unviable, leading to M&A.

This result can be shown to hold for incremental trade liberalization as follows. Suppose a Foreign firm is contemplating acquiring a Home firm. Trade liberalization increases the Home firm's profits on its initial exports, which makes it a more expensive takeover target, thereby making cross-border M&As less likely. However, trade liberalization also increases the Foreign firm's profits from exporting and reduces both firms' profits in their home markets, both of which make cross-border M&As more likely. In the neighborhood of autarky, the first effect does not arise. Therefore starting from a high level of trade barriers, a small amount of trade liberalization raises the possibility of cross-border M&As.

Thus both export-platform FDI and M&A type FDI could potentially explain increased FDI in the EU countries concomitant with trade liberalization. However, there is another possible explanation: this result could also arise if FDI is of the vertical type.

Vertical FDI Vertical FDI refers to the case when a multinational firm breaks its production process into parts and carries out different parts in different countries. A standard modeling approach is to divide the production process into two stages: headquarters stage and production stage where the former is carried in the home country while the latter is carried in a foreign country. The motivation for fragmentation of production comes from differential factor intensity of different stages of production and differences in factor prices. In the presence of differences in factor prices, minimizing cost requires each stage of production be undertaken in the country where the cost of doing so is the lowest. With vertical FDI, the goods produced in a foreign country have to be sent back to the parent country and hence are subject to

the trading cost. Therefore there is a trade-off between a lower cost of production in a foreign country and a trading cost of shipping these goods back to the parent country. In this framework, higher trading costs in the form of tariffs and transportation cost would discourage FDI. Therefore while horizontal FDI is encouraged by high tariffs and transportation cost, vertical FDI is discouraged by it. Consequently EU trade liberalization could generate increased FDI of the vertical type.

Outsourcing versus Vertical FDI Even if it is cheaper to conduct a part of the production process in a foreign country because of lower costs of production, the question remains as to whether this activity should be done within the firm boundary (FDI) or at arm's length (outsourcing). Analogous to the choice between licensing and FDI for a horizontal multinational, a recent literature has emerged that focuses on the choice between outsourcing and FDI for a vertical multinational. The key determinant of this choice again is the cost due to informational asymmetry and contract enforcement. In general, a high contracting cost makes FDI more attractive compared with outsourcing (see Helpman 2006 for a recent survey of this literature). Therefore, while a reduction in tariffs and transportation cost increases vertical FDI, a decrease in contracting cost favors outsourcing over FDI.

Other Trade Costs and FDI In addition to the types of trade costs discussed earlier, there are several other trading costs that affect both horizontal and vertical FDI. A legal system protecting the property rights of foreign investors reduces the risk of expropriation and hence encourages foreign investment. A strong intellectual property rights regime on the other hand may reduce FDI and increase licensing or outsourcing by reducing opportunistic behavior in arm's length transaction.

Another type of trade cost that could affect FDI is the cost associated with exchange rate volatility. Since multinational firms repatriate their profits from their activities in the host countries to their home countries, exchange rate fluctuations affect their profits in their home currency. Theoretical work on the relationship between exchange rate volatility and FDI

identifies offsetting effects. On the one hand, volatility increases the value of having production facilities in both the home country and the host country, allowing a multinational firm to shift production between the two plants depending on input costs. On the other hand, since FDI involves sunk costs to be incurred in the currency of the host country, greater volatility in exchange rate increases the option value of waiting for a multinational firm leading to reduced FDI. There is empirical evidence showing that exchange rate volatility reduces entry by multinational firms contemplating investment in the United States.

Policy Implications To the extent that FDI raises welfare in the host country, there would be scope for intervention to attract FDI. However, as discussed above, the relationship between trade costs and FDI is not straightforward. It depends on the type of trade cost and the type of FDI. While high tariffs and transportation cost may promote horizontal FDI of some types, they may discourage export-platform FDI and FDI through M&As. There is evidence of tariff-jumping FDI by Japanese firms in the automobile sector in the United States in the 1980s in response to a threat of higher tariffs. However, threatening a higher tariff to attract FDI is questionable strategy.

Tariffs and transportation cost usually reduce vertical FDI. If the good is assembled in the host country by a multinational subsidiary and sent back to the parent, then a higher tariff in the parent country will discourage this type of FDI. In this case, a higher tariff in the host country will also discourage FDI because usually some inputs are shipped from the parent country to the host country for assembly or further processing. If these inputs incur tariffs, then FDI would be discouraged. Therefore in the case of vertical FDI the policy implication would be for the host country to reduce tariffs to encourage FDI.

High contracting costs promote FDI at the expense of licensing in the horizontal case and at the expense of outsourcing in the vertical case. Therefore a reduction in the contracting cost is likely to reduce FDI and increase licensing and outsourcing. However, one should not jump to any policy conclusions here, because it is not clear whether FDI leads to greater welfare gains compared with licensing or outsourcing.

More important, the welfare implications of a policy of increasing trading cost to attract FDI is ambiguous even if FDI is welfare improving because an increase in trading cost has a direct negative effect on welfare that must be offset by the indirect positive effect arising from increased FDI.

In sum, the relationship between trade costs and FDI depends on the nature of FDI. Any view that trade costs always promote FDI is simplistic as it does not hold true for vertical FDI or even all types of horizontal FDI.

See also location theory; multinational enterprises

FURTHER READING

Anderson, James E., and Eric van Wincoop. 2004. "Trade Costs." *Journal of Economic Literature* 42 (3): 691 751. A survey on various types of trade costs and their importance for international trade.

Brainard, S. Lael. 1997. "An Empirical Assessment of the Proximity Concentration Trade off between Multinational Sales and Trade." *American Economic Review* 87 (4): 520 44. Empirical evidence on proximity concentration trade off using data on U.S. multinationals.

Dunning, John H. 1973. "The Determinants of International Production." *Oxford Economic Papers* 25 (3): 289 336. Pioneering work outlining the ownership location internalization framework to study multinationals.

Ekholm, Karolina, Rikard Forslid, and James R. Markusen. 2007. "Export Platform Foreign Direct Investment." *Journal of the European Economic Association* 5 (4): 776 95. Builds a theoretical model of export platform FDI.

Helpman, Elhanan. 2006. "Trade, FDI, and the Organization of Firms." *Journal of Economic Literature* 44 (3): 589 630. A survey of the recent literature on trade and FDI focusing on the organizational features such as sourcing strategies of firms.

Markusen, James R. 2002. *Multinational Firms and the Theory of International Trade.* Cambridge, MA: MIT Press. Unified view of trade and FDI resulting from the interaction of scale economies, trade costs, factor endowments, and imperfect competition.

Neary, J. Peter. 2007. "Trade Costs and Foreign Direct Investment." Mimeo. Oxford: Oxford University.

Explains how trade liberalization could increase M&A and export platform FDI.

Ranjan, Priya. 2006. "Preferential Trade Agreements, Multinational Enterprises, and Welfare." *Canadian Journal of Economics* 39 (2): 493 515. Explains how preferential trading agreements affect FDI.

Russ, Katheryn N. 2007. "The Endogeneity of the Exchange Rate as a Determinant of FDI: A Model of Entry and Multinational Firms." *Journal of International Economics* 71 (2): 344 72. Studies the impact of exchange rate volatility on FDI.

UNCTAD. 2000. *World Investment Report: Cross Border Mergers and Acquisitions and Development.* Geneva: United Nations. Provides data on M&As.

PRIYA RANJAN

■ trade facilitation

The expansion of world trade in the late 20th and early 21st centuries has been driven, in large part, by the changing nature of production and increased competition in international commerce. Another important factor has been the periodic rounds of successful multilateral trade negotiations. Talks at the World Trade Organization (WTO) have led to a considerable reduction in tariffs on goods crossing national borders. As the role of traditional trade barriers gradually vanishes, the focus of trade policy has shifted to lowering the remaining nontariff barriers to trade, including trade facilitation.

Trade facilitation involves a wide range of activities centered on lowering trade transaction costs for firms in global commerce. These costs include the price of moving freight from the factory to final destinations. Firms must manage border clearance procedures and pay trade services fees, among many other steps after goods and services are produced. As such, trade facilitation involves much more than trucking goods across national borders or shipping a package by sea transport.

Since trade facilitation can encompass a wide variety of measures, there is no single definition of the term. One might distinguish, however, between trade facilitation in a narrow or broad context.

Multilateral trade rules reflected in the three GATT articles V (freedom of transit), VIII (fees and formalities), and X (publication and administration of trade regulation) constitute a relatively narrow definition of trade facilitation. These articles call on WTO members to harmonize and simplify procedures related to the international exchange of goods and services.

Modern commerce involves a much broader scope of activities affecting trade facilitation not directly covered by WTO rules. This includes domestic policies, such as sanitary and phytosanitary standards, regulation of services, and use of information technology to lower trade costs. This wider definition takes into consideration the fact that "inside-the-border" issues such as the transparency of domestic regulations, institutional policies on foreign ownership of companies, and legal frameworks all affect a country's investment climate and transaction costs. In addition, the potential for efficiency savings due to electronic transmission of trade documentation and automation of customs operations suggest that global information technology infrastructure also drives trade facilitation.

The policy focus on these issues reflects the fact that trade can be a powerful engine for accelerating economic growth, job creation, and poverty reduction. Success in export markets for developed and developing country firms, therefore, is increasingly affected by the ability of countries to support an environment that promotes efficient and low-cost trade services and logistics.

Cross-Country Comparisons and Performance There is an increasing body of empirical evidence about the impact of trade facilitation on export competitiveness and growth. Studies reviewed by the Organisation for Economic Co-operation and Development (OECD 2002) indicate that trade transactions costs can amount to as much as 15 percent of the value of traded goods globally. A subsequent OECD study (2003) found trade transactions costs to be higher on agricultural and food products, fish, and forest and wood products (since these products are subject to additional border procedures due to sanitary and phytosanitary requirements). These are

products for which many developing countries have an advantage. The same study also reported that small and medium-size enterprises suffer most from poor trade-related practices and poorer developing countries have a larger share of such enterprises. A number of other studies have illustrated the specifics of why trade facilitation matters and specific sources of trade costs. One study, for example, found that barriers to export performance in Africa are closely related to firm characteristics and policies that raise trade costs. This includes nontransparent customs laws and administration. Much less evidence, in comparison, was found that transportation infrastructure had a significant impact on export performance (Clarke 2005).

The World Bank's *Doing Business* report (2007) documents the wide range of reform needed in developing countries to lower trade costs. The report outlines procedural requirements for importing and exporting a standardized cargo of goods in 155 countries. While the total cost to import (in U.S. dollars per container) was $842 on average in high-income countries, it was $1,960 in low-income countries. Typical regulations in low-income countries required 13 documents from domestic regulatory agencies, as compared with 6 signatures in high-income, 9 in upper-middle-income, and 10 in lower-middle-income countries. On average it still costs almost two and a half times in expenses and requires more than twice as many documents and four times as many signatures to trade in a poor country as it does in rich countries.

The *Doing Business* report also provides concrete examples of efficiency savings made possible through trade facilitation reforms. Much of these relate to addressing regulatory reform and other steps that in contrast to hard infrastructure constitute the major part of why engaging in trade takes longer in developing countries. Progress in reducing costs, however, has been made. For example, Guatemala with the support of the Inter-American Development Bank changed to an electronic system for export authorization in 2000. Within four years the time for authorization of export documents dropped from one day to around three minutes. Tunisia has also introduced an automated system that provides a one-stop trade documentation-processing platform. Due to this innovation the processing time for trade documentation was reduced from 18 to 7 days, which probably has led to substantial productivity gains according to the United Nations Economic Commission for Africa.

The Gains to Trade from Cutting Costs and Red Tape What about the gains to trade with an ambitious reform agenda? There are important gains to lowering trade costs and facilitating trade with some variation across countries and sectors (Francois et al. 2005; OECD 2003). Several studies have focused on the impact of trade facilitation on the microlevel (e.g., Nomura Research Institute 2004; Keen 2003; OECD 2005). They observe that in some developing countries inefficient trade regulations, documents, and procedures are hindering firms' participation in export markets. The studies find large potential benefits from streamlining trade regulations and hence cutting the costs of exporting.

Analysis by the World Bank suggests that raising global capacity in trade facilitation (port efficiency, customs, regulatory transparency, and information technology used in trade transactions) halfway to the global average would increase world trade by $377 billion (Wilson et al. 2005). This is an increase of about 9.7 percent in global trade with the majority of the gains due to domestic reform and capacity building. About $107 billion of the total gain comes from the improvement in port efficiency and about $33 billion results from the improvement in customs environment. The gain from the improvement in regulatory environment is $83 billion. The largest gain ($154 billion) comes from an improvement in information technology infrastructure, which is increasingly applied all across the trade value and logistics chain. The important work ahead to lower trade costs, however, is outside the scope of trade policy and squarely centered within domestic reform agendas, including transportation, telecommunications, and other services that affect trade transactions costs. There is also a key role to be played by private sector led policy reform.

Moving Ahead: Cutting Trade Costs Talks on revisions to the WTO rules on trade facilitation continued on the agenda in the Doha Development Agenda negotiations at the end of 2007. The outlines of a possible agreement included, among other provisions: (1) new obligations to promote electronic distribution and transmission of government trade regulations on imports and exports, (2) standardization of certain basic fees for imports, and (3) stronger rules to help ensure freedom of transit for goods crossing national borders. The WTO Secretariat tabulation of proposals from August 11, 2006, includes consideration of more ambitious changes to current obligations, such as establishment of a single window for exporters and importers at customs or accepting copies of documents for import and export in lieu of originals. Since 2002 there have also been talks about explicit ties between a package of technical assistance and development aid as part of any final agreement. Developing countries participated very actively in these negotiations and before the suspension of the Doha Round in late 2006 considerable progress had been achieved (World Bank 2006; Neufeld 2007).

Changes to the multilateral trade rules noted above can help in advancing transparent and predictable trading procedures for exporters. These rules have not been revised in more than 50 years, and crafting an achievable and focused next step in the Doha Agenda will be helpful to traders. This is especially true for small and medium-sized firms that lack resources and networks to overcome some of the more complex border rules now in place. Meeting trade facilitation goals in a broader and deeper context, however, will require action well outside and beyond the WTO framework. More effective delivery of development aid and technical assistance especially in projects to advance regulatory reform and support public agencies that promulgate and administer trade regulations is needed. A better understanding is required of conditions under which developing countries are ready to receive such assistance so that aid is effectively delivered.

The Agenda Ahead At the most basic level, better data, analysis, and indicators of performance are needed to inform discussions moving forward on how trade costs and facilitation measures affect global commerce. Progress in reaching development goals in trade can be made only with accurate data and analysis to drive policy choice and action. Taking advantage of the energy behind reaching the Millennium Development Goals (MDGs) and building on the data sets in the Global Monitoring Reports, which track progress in reaching poverty reduction goals, among other reports, makes significant sense. In sum, the data gathered to date illustrate the fact that the nontariff agenda in trade is increasingly important to economic development.

See also Doha Round; nontariff measures; sanitary and phytosanitary measures; technical barriers to trade; trade-related capacity building; World Trade Organization

FURTHER READING

Clarke, George. 2005. "Beyond Tariffs and Quotas: Why Don't African Manufacturers Export More?" World Bank Policy Research Working Paper No. 3617 (June). Washington, DC: World Bank. An empirical investi gation of how weak customs regimes impede the export performance of manufacturing firms in African coun tries.

Francois, J., H. van Meijl, and F. van Tongeren. 2005. "Trade Liberalization in the Doha Development Round." CEPR Discussion Papers No. 4032. London: Center for Economic Policy Research. Provides esti mates of possible welfare benefits from a successful completion of multilateral trade negotiations, including those on trade facilitation.

Keen, M. 2003. "The Future of Fiscal Frontiers and the Modernization of Customs Administration." In *Chang ing Customs: Challenges and Strategies for the Reform of Customs Administration*, edited by M. Keen. Wash ington, DC: IMF. Available online at http://www.imf .org/external/pubs/nft/2003/customs/. Outlines key rea sons why modernization of customs administration is crucial and describes important aspects of customs re form.

McLinden, Gerard. 2006. "Trade Facilitation: Progress and Prospects for the Doha Negotiations." In *Trade, Doha, and Development: A Window into the Issues*, edited by

R. Newfarmer. Washington, DC: World Bank, 175 85. An examination of the progress in the WTO negotia tions and the role of new WTO obligations to imple ment trade reforms in developing countries.

Neufeld, N. 2007. "WTO Negotiations on Trade Facil itation Lessons for the Future? New Perspectives for and from the Developing World." In *Developing Countries and the WTO: Policy Approaches*, edited by G. Sampson and W. Bradnee Chambers. Tokyo: United Nations University Institute of Advanced Studies. A summary of recent multilateral trade negotiations on trade facilitation with a particular emphasis on the role of developing countries.

Nomura Research Institute. 2004. "Study on the Economic Effects Resulting from Reduction of Lead Time for Import Procedures." Customs and Tariff Bureau, Ministry of Finance of Japan, March 2004. An empirical review of how Japan was able to reduce the lead time of import procedures and the economic impact of these changes.

OECD. 2002. *Business Benefits of Trade Facilitation*. Paris: OECD (November). Provides an assessment of the business benefits of trade facilitation based on a survey of studies.

. 2003. *Quantitative Assessment of the Benefits of Trade Facilitation.* Paris: OECD. A detailed empirical exploration of trade facilitation variables and their pos sible impact on trade.

. 2005. "Trade Facilitation Reforms in the Service of Development: Country Case Studies." TD/TD/WP (2004)4/FINAL. Paris: OECD. Presents country stud ies (Mozambique, Angola, Pakistan, and Peru) of cus toms reform.

Wilson, John S. 2005. "Trade Facilitation and Economic Development." In The World Trade Brief. For the Fourth WTO Ministerial Conference, Hong Kong, 13 18 December. London: Agenda Publishing, 46 48. Provides a quick literature review on the topic, including important quantitative results.

Wilson, John S., Catherine L. Mann, and Tsunehiro Otsuki. 2005. "Assessing the Benefits of Trade Facil itation: A Global Perspective." *The World Economy* 28 (6): 841 71. An empirical investigation of the relationship between broad indicators of trade facilita tion performance, trade flows, and benefits of reform.

World Bank. 2006. *Doing Business in 2007*. Washington, DC: World Bank. This database records regulatory and others costs of doing business in 175 economies. Ac cessible at http://www.doingbusiness.org/.

JOHN S. WILSON

■ trade in services

Since the conventional definition of trade where a product crosses the frontier would miss out on a whole range of international services transactions, it is now customary to take a broad view of ser vices trade to include all modes of conducting in ternational transactions. Services trade, according to this wide definition, is an increasingly important part of global commerce. Advances in information and telecommunication technologies have expanded the scope of services that can be traded across borders, from accounting to education. Many countries now allow foreign investment in newly privatized and competitive markets for services, such as tele communications, transportation, and finance. More and more people are traveling abroad to con sume tourism, education, and medical services, and to supply services ranging from construction to software development.

The Emerging Pattern of Trade Rigorous anal ysis of patterns of trade in services has been inhibited by the paucity of data. The only services trade sta tistics available on a global basis are the International Monetary Fund's balance of payments (BOP) sta tistics, which register transactions between residents and nonresidents. These statistics suggest that ser vices have been among the fastest growing compo nents of world trade, growing by more than 15 per cent per year since 1980. The value of global services exports is now close to $3 trillion, representing more than one-fifth of aggregate world trade in goods and services. This value is certainly understated, because it does not include sales through an established presence, that is, via FDI and by individuals who stay longer than a year, both of which are treated as resi dents by the BOP statistics. Today more than half of annual world FDI flows are in services, and the value

of sales abroad by foreign affiliates of U.S. service firms is estimated to be 3.5 times greater than their cross-border exports.

The United States and the European Union lead the world in a broad spectrum of services trade telecommunications, transportation, finance, retail trade, and a range of business services. Together they account for more than 60 percent of world services exports. China and India jointly account for only 3 percent. But several developing countries are emerging as the most dynamic services exporters. Over the last decade, while the exports of the United States have grown at around 10 percent per year, the exports of countries such as India, China, and Brazil have grown at rates well above 15 percent per year. Some developing countries are even beginning to invest abroad to provide services for example, Malaysia in environmental services, South Africa in telecommunications and retail, and India in software but many more are supplying services via cross-border sales (e.g., data processing), tourism services for visiting foreign consumers, and the movement abroad of individual services providers (e.g., professional services or construction workers).

What Determines the Pattern of Services Trade? The two major explanations for trade between countries, comparative advantage and gains from specialization arising from increasing returns to scale or agglomeration effects, are equally relevant to services trade. Moreover both explanations apply not only to cross-border trade but also to other modes of trade, including commercial presence and the movement of people. Examples of trade based on comparative advantage include call centers in India that provide customer contact services for U.S. firms; nannies from the Philippines who move to Canada temporarily to provide child care services; and Europeans who travel to Peru for a week in the jungle as part of an ecotourism package. The trade in both child care and call center services is driven by differences in labor costs across countries; and the Amazon has unique attributes that are not available at home to the European tourists.

Although differences between countries are one of the major explanations for services trade, particularly between countries with very different income levels, they cannot account for all trade. Much of the world's trade occurs between high-income countries, and much of the trade between similar countries is in similar services. For example, European banks operate in the United States and U.S. banks in Europe. Canadian engineers work on projects in the United States, and U.S. engineers work on projects in Canada. A labor market example helps explain how services trade can emerge between similar countries. Two students starting university are equally bright and talented. At this point their productive capacities may seem indistinguishable. However, suppose one chooses to study medicine and the other chooses to study engineering. If we revisit these same students 10 years later, they will have very different skills and they can trade with each other via the labor market, with the doctor selling medical services and the engineer selling engineering services. This example illustrates the genesis of a large part of trade today in services ranging from finance and communications to transport and education all services where there are large economies of scale and consumers desire variety.

What Restricts Services Trade? In spite of recent liberalization, services trade still tends to be quite restricted. First, many types of services were publicly provided and a significant number continue to be produced by state-owned or regulated monopolies. The traditional rationale was the existence of natural monopoly in services such as telecommunications and transportation, and a social concern for equitable access in services such as health care and education. Second, entry and operation in a range of services, from financial to professional, are regulated on the grounds that it is particularly difficult for consumers to judge the quality of these "invisible" services before they are consumed.

In contrast to goods, relatively few services are subject to simple discriminatory taxes on trade. Instead barriers to trade in services arise from domestic regulations that often serve the dual purpose of responding to market failures (such as ensuring quality standards for medical practitioners) and protecting

local suppliers from foreign competition. This means that identifying and measuring trade barriers in the service sector is very complex. It also means that simple rules for trade liberalization that have worked for goods trade (such as reducing all tariffs by 30 percent) are not available as an option for service trade liberalization. Instead service trade liberalization is organized around the notion of nondiscrimination and is often linked with domestic regulatory reform.

Even though tariffs are rare, quotas are pervasive. On cross-border trade, they are most evident in the transportation sectors. Foreign providers are either completely shut out of certain segments, such as cabotage (i.e., transport between locations within a country), or only provided limited access, as in international transportation. On commercial presence, quotas are imposed on the number of foreign suppliers who are allowed to enter sectors such as telecommunications and banking, or on the extent of foreign ownership in individual enterprises. Quotas are most stringent in the case of movement of service-providing personnel and affect trade not only in professional services but also in a variety of labor-intensive services.

What Are the Consequences of Liberalization?
Economic theory suggests that a country gains from the import of services, irrespective of the mode of transaction, if the terms at which international transactions take place are more favorable than those available on the domestic market. The limited empirical evidence suggests that liberalization of trade in services, accompanied by the reform of complementary policies, can produce substantial gains.

Removing barriers to trade in services in a particular sector is likely to lead to lower prices, improved quality, and greater variety. As in the case of trade in goods, restrictions on trade in services reduce welfare because they create a wedge between domestic and foreign prices, leading to a loss to consumers that is greater than the increase in producer surplus and government revenue. Furthermore, since many services are inputs into production, the inefficient supply of such services prevents the realization

of significant gains in productivity for downstream firms. As countries reduce tariffs and other barriers to trade in goods, effective rates of protection for manufacturing industries may become negative if they continue to be confronted with input prices that are higher than they would be if services markets were competitive.

Several empirical studies provide evidence of gains (reviewed in Deardorff and Stern 2007). Estimates of benefits vary for individual countries depending on the initial levels of protection and the assumed reduction in barriers. Developed countries tend to gain more in absolute terms — which is not surprising given the relative size of their economies — but developing countries also see significant increases in their gross domestic product (GDP). One model predicts gains of between 1.6 percent of GDP (for India) to 4.2 percent of GDP (for Thailand) if tariff equivalents of protection were cut by one-third in all countries. The gains from liberalizing services are usually estimated to be substantially greater than those from liberalizing trade in goods, because current levels of protection are higher and because liberalization would also create spillover benefits from the required movement of capital and labor. Finally there is also some evidence — relatively strong for the financial sector and less strong but nevertheless statistically significant for the telecommunications sector — that openness in services significantly influences long-run growth performance.

These results are particularly striking because they are derived from models that do not fully allow for the temporary movement of individual service suppliers — potentially a major source of gain. Such movement offers arguably the neatest solution to the dilemma of how international migration is best managed, enabling the realization of gains from trade while averting social and political costs in host countries and brain drain from poor countries. Recent research finds that if members of the Organisation for Economic Co-operation and Development (OECD) were to allow temporary access to foreign services providers equal to just 3 percent of their labor force, the global gains would be more than $150 billion — more than the gains

from the complete liberalization of all trade in goods (Winters et al. 2003). Both developed and developing countries would share in these gains, and they would be largest if both high-skilled mobility and low-skilled mobility were permitted.

How Is Liberalization Best Implemented? It would be wrong to infer that these gains can be realized by a mechanical opening up of services markets. A flawed reform program can undermine the benefits of liberalization.

For example, if privatization of state monopolies is conducted without concern to creating conditions of competition, the result may be merely transfers of monopoly rents to private owners (possibly foreigners). Similarly, if increased entry into financial sectors is not accompanied by adequate prudential supervision and full competition, the result may be insider lending and poor investment decisions. Also, if policies to ensure universal service are not put in place, liberalization need not improve access to essential services for the poor. Managing reforms of services markets therefore requires integrating trade opening with a careful combination of competition and regulation.

Sustaining openness may also require action to alleviate adjustment costs, even though the available evidence suggests that openness to services trade has not hurt overall employment. For example, in the United States, which in recent years has seen a significant growth in services outsourcing, private employment in services has actually expanded (in contrast to manufacturing) by more than a million jobs. The structure of employment has changed but more because of new technology than trade. Between 1999 and 2003, it is estimated that in the United States more than 70,000 computer programmers lost their jobs, whereas more than 115,000 higher-paid computer software engineers obtained jobs (Kirkegaard 2004). As in the case of goods trade, assisting dislocated workers is imperative if the gains from new trade opportunities and technological progress are to be widely shared. But this assistance is best provided through improved income support, retraining, and health insurance, rather than through restrictions on technology and trade.

To conclude, services trade continues to grow rapidly thanks to technological change and policy reform. Industrial countries dominate international transactions, but several developing countries are among the most dynamic participants. Trade has already produced significant benefits but the potential gains are even larger. There is a strong economic case for eliminating remaining barriers to trade, in particular the restrictions on foreign investment and temporary migration. But successful liberalization needs complementary policy action to improve the regulatory environment, to alleviate adjustment costs, and to ensure that the benefits of openness are widely shared.

See also comparative advantage; General Agreement on Trade in Services (GATS); World Trade Organization

FURTHER READING

Deardorff, Alan V., and Robert M. Stern. 2007. "Empirical Analysis of Barriers to International Services Transactions and the Consequences of Liberalization." In *Handbook of Services Trade*, edited by Aaditya Mattoo, Robert M. Stern, and Gianni Zanini. Oxford: Oxford University Press, 169 220. A review of attempts to quantify the magnitude and implications of trade barriers.

Hindley, Brian, and Alasdair Smith. 1984. "Comparative Advantage and Trade in Services." *World Economy* 7 (4): 369 90. A persuasive article arguing that the economics of services trade is not fundamentally different from the economics of goods trade.

Hoekman, Bernard. 2006. "Liberalizing Trade in Services: A Survey." World Bank Policy Research Working Paper No. 4030. Washington, DC: World Bank. An up to date review of the existing literature on services trade.

Kirkegaard, Jacob. 2004. "Outsourcing: Stains on the White Collar." Mimeo. Washington, DC: Institute for International Economics.

Mattoo, Aaditya, Robert M. Stern, and Gianni Zanini, eds. 2007. *Handbook of Services Trade*. Oxford: Oxford

University Press. A comprehensive volume that includes chapters on the economics of services trade, measuring services trade, and empirical analysis of barriers to trade as well as on specific sectors.

Winters, L. A., T. L. Walmsley, Z. K. Wang, and R. Grynberg. 2003. "Liberalizing Temporary Movement of Natural Persons: An Agenda for the Development Round." *World Economy* 26 (8): 1137 61. Estimates the potential gains from the OECD countries allowing in creased immigration.

AADITYA MATTOO

■ Trade Policy Review Mechanism

The Trade Policy Review Mechanism (TPRM) involves periodic reviews of the trade and trade-related policies, practices, and measures of each member of the World Trade Organization (WTO). The reviews take place in the WTO's Trade Policy Review Body (TPRB), a senior body composed of representatives of all WTO members. The WTO's four largest trading entities (the United States, the European Community, China, and Japan) are reviewed every two years; the next sixteen are reviewed every four years; and other members are reviewed every six years, although a longer period may be allowed for least-developed countries. Each review is based on a report prepared by the member(s) under review and on a report drawn up by the WTO Secretariat, in its Trade Policies Review Division (TPRD), on its own responsibility. Both reports are published.

The TPRM's objective is to achieve greater transparency in, and understanding of, members' trade policies, practices, and measures. It permits the evaluation of trade and related policies and measures, including those that do not necessarily contravene, or indeed are not subject to, WTO obligations.

Transparency entails four key elements: (1) a description of the nature of policies and measures; (2) their rationale or objectives; (3) their costs (in terms of expenditures or taxes forgone); and (4) an economic evaluation of their effectiveness of policies and measures (relative to alternatives) in achieving their objectives. Accordingly, the TPRM enables the regular collective appreciation and economic evaluation of a full range of individual members' trade policies and practices, their consistency with the broad principles of nondiscrimination and predictability that underlie the WTO, and their impact on the functioning of the multilateral trading system.

By reviewing broad macroeconomic and structural policies, trade policy reviews (TPRs) attempt to place trade and trade-related policies in their broader policy setting, thereby assessing the coherence of these policies in achieving their objectives. In this regard, TPRs are being seen increasingly by many countries as a useful form of technical assistance.

The TPRM is not intended to serve as a basis for the enforcement of obligations under WTO agreements or for dispute-settlement purposes. Nor is it intended to impose new rules on members. Rather, one purpose is to contribute, essentially by moral suasion, to improved adherence by all members to disciplines and commitments agreed to under the WTO and thus to a smoother functioning of the multilateral trading system.

Trends and Lessons from Reviews Reviews conducted since the establishment of the TPRM have brought to light the benefits countries have obtained through their integration into the world trading system, while also highlighting the difficulties encountered by many (often developing) countries in this process. By and large, countries have become integrated in the world economy through unilateral measures and the multilateral trading system rather than through initiatives they have taken in regional agreements. Thus in numerous countries, developed and developing alike, unilateral liberalization of trade and investment policies, as well as deregulation programs, have taken place. Such reforms are often part of comprehensive integrated policy packages that include macroeconomic and other structural reform measures.

More generally, the reviews have shown that in most trade regimes tariffs have largely replaced quantitative restrictions and other nontariff measures as the main form of protection, although the process remains incomplete in some cases. Many tariff structures have been or are being rationalized and average tariffs have been reduced.

Although there have been reversals of trade liberalization in some countries (in the form of tariff increases, new surcharges on imports, or the re-imposition of quantitative restrictions), in most cases these setbacks have been limited, rarely offsetting the earlier liberalization measures. The reviews have also shown that, in general, the experience in trade reform in less-developed countries has lagged behind that of developing countries.

Members that have undergone a review have repeatedly testified to the value of the experience for the development of their own trade policies. Past experience shows that through a review members engage in a process of self-examination in which they reflect both on their own policies and on their participation in the multilateral trading system and their engagement with the WTO. Many countries have started internal reviews of their trade policies on a yearly basis.

Inconsistencies identified in trade regimes have made members more aware of the necessity of having consistent trade policies, formulated and implemented at national levels, with the involvement of all the relevant departments. Thus in the process of reviews, members also increase their awareness of the economic role of customs tariffs, broadening their perception of tariffs as a fiscal policy instrument only. Many countries have also been led, in consequence of their reviews, to take steps to address some of the concerns likely to be expressed by other WTO members; as a result many, particularly less-developed, countries have improved compliance to their WTO commitments in areas such as notification. Throughout the process, members develop confidence in the TPRD professionals involved in their reviews and use them as de facto inquiry/focal points on WTO issues. More generally, the reviews help build knowledge in, and understanding of, the WTO. Also,

members that are in a process of trade policy reform have an opportunity to present the challenges, the process, and its results to a world audience. Discussions take place in the TPRD and with WTO economists that encourage trade policymaking to take directions foreseen in the WTO agreements and contribute to a member's greater integration into the multilateral trading system. In this context countries are provided with a framework for the formulation, implementation, and assessment of their trade policies and with an instrument to facilitate interministerial consultations for these purposes. The TPRM now also increasingly undertakes group reviews, for example of the Southern African Customs Union, the East African Community, and the Organization of East Caribbean States. This can foster regional integration that complements the WTO system.

Reviews also allow future technical-assistance priorities to be identified more clearly. In this respect, each Secretariat report for a review of a less-developed country now includes a section on capacity-building and technical-assistance needs; there is now also close cooperation with the WTO's Institute for Training and Technical Cooperation such that a needs assessment can be established and elaborated in the member's own report. The needs are invariably discussed during the review and have subsequently often been endorsed by the TPRB, thus forming the basis for a well-articulated future program of technical-assistance activities. In this context also, the process of reviewing less-developed countries now responds more systematically to technical-assistance needs. The review process for a less-developed country, and indeed for many developing countries, includes a two- to three-day seminar for its officials on the WTO and, in particular, the trade policy review exercise and the role of trade in economic policy.

Coverage of the TPRM Table 1 provides a list of WTO members reviewed in the period 1989 2007. In all, the TPRM conducted 248 reviews by the end of 2007. Increasing emphasis is given to less-developed countries, both because such members ask for reviews as an important element in their own

Table 1
Trade policy reviews: WTO members reviewed, 1989 2007

Europe/Middle East	Asia/Pacific	Africa	America
Austria[c]	Australia (5)	Angola[b]	Argentina (3)
Bahrain (2)	Bangladesh[b] (3)	Benin[b,h] (2)	Antigua and Barbuda[f] (2)
Bulgaria	Brunei Darussalam	Botswana[e] (2)	Barbados
Cyprus[c]	China	Burkina Faso[b,h] (2)	Belize
Czech Republic[c] (2)	Chinese Taipei	Burundi[b]	Bolivia (3)
European Communities (8)	Fiji	Cameroon[h] (3)	Brazil (4)
Finland[c]	Hong Kong, China (5)	Central African Republic[a,b]	Canada (8)
Hungary[c] (2)	India (4)	Chad[a,b]	Chile (3)
Iceland (3)	Indonesia (5)	Congo[a]	Colombia (3)
Israel (3)	Japan (8)	Côte d'Ivoire	Costa Rica (3)
Liechtenstein[d] (2)	Korea, Rep. of (4)	Djibouti[b]	Dominica[f] (2)
Norway (4)	Kyrgyz Republic	Egypt (3)	Dominican Republic (2)
Poland[c] (2)	Macau, China (3)	Gabon[h] (2)	Ecuador
Qatar	Malaysia (4)	Gambia[b]	El Salvador (2)
Romania (3)	Maldives[b]	Ghana (2)	Grenada[f] (2)
Slovak Republic[c] (2)	Mongolia	Guinea[b] (2)	Guatemala
Slovenia[c]	New Zealand (3)	Kenya[g] (3)	Guyana
Sweden[c] (2)	Pakistan (2)	Lesotho[b,e] (2)	Haiti[b]
Switzerland (4)	Papua New Guinea	Madagascar[b]	Honduras
Turkey (4)	Philippines (3)	Malawi[b]	Jamaica (2)
United Arab Emirates	Singapore (4)	Mali[b,h] (2)	Mexico (3)
	Solomon Islands[b]	Mauritania[b]	Nicaragua (2)
	Sri Lanka (2)	Mauritius (2)	Panama[a]
	Thailand (5)	Morocco (3)	Paraguay (2)
		Mozambique[b]	Peru (3)
		Namibia[e] (2)	St.Kitts and Nevis[f] (2)
		Niger[b,h]	St.Lucia[f] (2)
		Nigeria (3)	St.Vincent & Grenadines[f] (2)
		Rwanda[b]	Suriname
		Senegal[b,h] (2)	Trinidad and Tobago (2)
		Sierra Leone[b]	United States (8)
		South Africa[e] (3)	Uruguay (3)
		Swaziland[e] (2)	Venezuela (2)
		Tanzania[b,g] (2)	
		Togo[b] (2)	
		Tunisia (2)	
		Uganda[b,g] (3)	
		Zambia[b] (2)	
		Zimbabwe	
37 members (48 reviews)	**24 members (69 reviews)**	**39 members (62 reviews)**	**33 members (69 reviews)**

Notes:

( ) = Number of reviews completed where this is greater than one.

Reviews conducted at end 2007 = 248.

WTO members reviewed = 133 out of 151 members.

Least developed WTO members reviewed = 27.

Share of world trade of WTO members reviewed (excluding significant double counting and intra EC trade) = around 97%.

[a]First review in 2007.

[b]Least developed member.

[c]Now included in European Communities (EC).

[d]Joint review with Switzerland (counted as two members, but one review for statistical purposes).

[e]Reviewed as member of the Southern African Customs Union (counted as five members, but one review).

[f]Reviewed as a member of the Organization of East Caribbean States (counted as six members, but one review).

[g]Reviewed as a member of the East African Community (counted as three members, but one review).

[h]Joint Review (two members, but one review for statistical purposes).

policymaking process and because the membership sees these reviews as essential to coming to a better understanding of the challenges that these members face and of how to respond to them.

WTO members have twice appraised the operation of the TPRM. In reports to WTO Ministerial Conferences they concluded that the mechanism, as the only multilateral, comprehensive evaluation of trade policies, functioned effectively, demonstrated a valuable public-good aspect, was a vehicle for members to reflect on their policies, served as an input into policy formulation, and highlighted general technical-assistance needs.

See also multilateralism; trade and economic development, international; trade-related capacity building; World Trade Organization

FURTHER READING

World Trade Organization. 1994. *The Results of the Uruguay Round of Trade Negotiations*. Annex 3. Geneva: WTO. Annex 3 provides the mandate of the TPRM.

——. 1995, 2005. *Rules of Procedure for Meetings of the Trade Policy Review Body*. WTO documents WT/TPR/6 and WT/TPR/&/Rev.1. Geneva: WTO. Sets out the procedures under which the TPRB conducts its work.

——. 1999, 2005. *Appraisal of the Operation of the Trade Policy Review Mechanism: Report to Ministers*. WTO documents WT/MIN(99)2 and WT/MIN(05)1. Geneva: WTO. Reaffirmed the importance that WTO members attach to the TPRM and found it to be functioning efficiently.

——. 2000 2006. *Trade Policy Review Mechanism: Report of the Trade Policy Review Body*. WTO documents WT/TPR/86, WT/TPR/101, WT/TPR/122, WT/TPR/140, WT/TPR/154, WT/TPR/173 and WT/TPR/192. Geneva: WTO. Provides details of the work of the TPRB in each of the years 2000 to 2006.

CLEMENS F. J. BOONEKAMP

■ trademarks

See Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS); intellectual property rights

■ trade-related capacity building

Trade-related capacity building (TRCB) helps developing and transition economies to benefit from trade liberalization and a growing volume of trade opportunities. The underlying philosophy of this form of foreign aid is based on the conviction that trade positively affects economic development and provides domestic clients access to a greater variety of goods and services at lower prices. The aim of TRCB is to improve information flow, skills, relationships, institutional arrangements, and infrastructure, which affect a developing country's trade performance. This form of development aid enhances organizational and human capital through training, conducting research and consultancy projects, learning-by-doing, networking, and establishment of appropriate facilities. It aims at improving the beneficiary country's export performance (including widening its range of exports and selling in a greater number of markets), attracting foreign investment to create jobs and trade, and ensuring that the country benefits from the institutions of international trade, especially the World Trade Organization (WTO).

Depending on the area covered, two types of TRCB may be distinguished: (1) capacity building addressing trade policy issues and (2) capacity building in the area of trade promotion. The latter aims at improving the international business performance of the beneficiary country's firms and trade support institutions. The former focused on trade policy and trade-related policy issues develops know-how, networks, and institutions with the objective of optimizing the beneficiary country's policies and procedures, as well as implementation and control.

Capacity building aimed at business promotion may be seen as a value chain disaggregated into two types of strategically relevant activities: (1) primary activities, which are essentially marketing related, and (2) support activities, which provide the inputs needed for primary activities to occur. Primary activities relate to export market analysis, quality improvement of export products, transfer of technology and management skills, partner search, institution building in the area of trade and foreign direct in-

vestment support institutions, and improvements in the performance of trade infrastructure. The support activities comprise the maintenance of business data networks; organization of seminars, workshops, and corporate internships; publication of the relevant training material; execution of business consulting projects; organization of buyer-seller meetings, fairs, and expositions; and the like.

The nature and scope of TRCB have been evolving in response to the larger issue coverage of the trading system, increased participation of developing countries in the WTO and regional trade agreements, increased focus on market orientation and governance, and greater complexity of trade rules and procedures. Government administration in developing economies has been particularly affected by these developments. In addition there has been a clear tendency in the developing world toward a proactive involvement in trade negotiations and an increased role of advocacy in domestic and international decision making. In such an environment, it has been widely felt that the relevant institutions are weak and qualified personnel not easily available. Especially in least-developed countries (LDCs), the complex WTO procedures and requirements necessitated institutional capacities and budgetary resources that could not be provided, and TRCB was needed to reduce the transaction costs of integration into the global market. There was also a growing need for TRCB in trade promotion resulting from the desire to intensify the role of markets and the private sector, and a growing international orientation in development strategies. International rules are meaningless if they are not used, and the rules' real users   the business firms   need managerial capacities and support by trade-facilitation agencies, chambers of commerce, and business associations in order to compete in global markets.

Actors in the TRCB process include donors and beneficiaries, providers, policymakers, numerous government departments and specialized development agencies, the business community, nongovernmental organizations (NGOs), and academia. Considerable resources are donated to multilateral, regional, and bilateral TRCB activities. The budget

of the TRCB projects of the U.S. Agency for International Development (USAID) alone   the leading supporter of TRCB   was close to US $700 million in fiscal year 2005, and it is estimated that more than US $2 billion per annum were spent on TRCB in the first years of this millennium. The other major donors include the European Union (EU) and the EU member states, such as the United Kingdom, Germany, and France, the Scandinavian countries, and Japan, Canada, and Australia. (A TRCB database established by the WTO jointly with the Organisation for Economic Co-operation and Development provides detailed information on TRCB projects.)

Most donors channel their support through governmental institutions (multilateral, regional, and national), NGOs, academia, or private firms. Among the beneficiaries are government departments and agencies, educational units, business associations, prioritized business firms, and other members of civil society. Among the leading providers of TRCB at the multilateral level are the World Bank and the International Monetary Fund (IMF), the WTO Secretariat, the United Nations Conference for Trade and Development (UNCTAD), the United Nations Development Program, the International Trade Center (ITC), and the World Intellectual Property Organization (WIPO). The WTO's assistance puts emphasis on trade policy rules, procedures, and rights and obligations of the WTO member states. The TRCB projects implemented by the World Bank, IMF, and UNCTAD emphasize the policy outcome and analysis of trade issues in the larger macroeconomic context. The ITC activities in TRCB are focused on issues of interest to the private sector, and WIPO concentrates on trade-related aspects of intellectual property. Regional governmental organizations conducting TRCB include the Organisation for Economic Co-operation and Development, the Asia Development Bank, the Secretariat of the Asia-Pacific Economic Cooperation (APEC) Forum, and so on.

At the national level many donor countries distinguish between the government departments in charge of formulating international development policies and the government-sponsored

implementing organizations. The former comprise ministries such as the Department for International Development in the United Kingdom, the Federal Ministry for Economic Cooperation and Development (Bundesministerium für Wirtschaftliche Zusammenarbeit und Entwicklung) in Germany, the Ministry of Foreign Affairs in Japan, and the Department for International Cooperation and Development (Coopération internationale et développement) in France. Among the national providers, the leading agencies include USAID (an independent federal government agency that receives overall foreign policy guidance from the Secretary of State), the Canadian International Development Agency, Deutsche Gesellschaft für Technische Zusammenarbeit GmbH, the Japan International Cooperation Agency, Le Groupe de l'Agence française de développement in France, and the Australian Agency for International Development.

The nongovernmental or mixed government/ private providers include a wide range of institutions such as the International Center for Trade and Sustainable Development (ICTSD), the Agency for International Trade Information and Cooperation, the North-South Institute, and the Center for Trade Policy and Law. An example of a nonprofit organization building trade law capacity in southern Africa is the Trade Law Centre for Southern Africa. The World Trade Institute, a joint venture between the University of Bern and the University of Neuchâtel in Switzerland, is a good example of another academic institution involved in TRCB with a special focus on international trade law. The role of such organizations is gaining importance with growing attention being paid to local ownership, efficiency, and diversity of TRCB services.

The lack of cooperation or coordination between the various providers is often quoted as one of the weaknesses of capacity building. Efforts for improvement included the launching of multiagency programs such as the Integrated Framework for Trade-Related Technical Assistance Programs to LDCs and the Joint Integrated Technical Assistance Program to selected LDCs and other African countries jointly conducted by ITC, UNCTAD, and WTO. It is expected that important "aid for trade" provisions of the Doha Development Round (WTO) will further reinforce the TRCB activities.

The TRCB multilateral activities in the beginning of the 21st century were dominated by concerns such as (1) assistance in the WTO accession negotiations, (2) developing country compliance with the WTO rules and institution building to reap the benefits of WTO membership, (3) promotion of developing country service industries through appropriate reforms, (4) trade in agricultural products, (5) trade facilitation, (6) trade-related infrastructure development, (7) trade and environment, (8) modernization of the financial system, (9) improvements in labor standards, (10) competition policy, (11) procedures of trade policymaking and governance, and (12) assistance in dispute settlements or bilateral consultations on trade disputes.

See also aid, international; technical barriers to trade; trade and economic development, international; World Bank; World Trade Organization

FURTHER READING

Finger, J. Michael, and Philip Schuler. 1999. "Im plementation of Uruguay Round Commitments: The Development Challenge." World Bank Policy Research Working Paper 2215. Washington, DC: World Bank. The paper discusses the post Uruguay Round chal lenges of trade related capacity building.

Kostecki, Michel. 2001. "Technical Assistance Services in Trade Policy: A Contribution to the Discussion on Capacity Building in the WTO." Geneva: International Center for Trade and Sustainable Development. Avail able at http://www.ictsd.org/pubs/respaper/TApaper 5 12 01.pdf. Comprehensive overview of trade related capacity building, emphasizing the role of international organizations.

Organisation for Economic Co operation and Develop ment (OECD). 2001. "The DAC Guidelines: Strengthe ning Trade Capacity for Development." Paris: OECD. A document that helped to shape developments in trade related capacity building.

Prowse, Susan. 2002. "The Role of International and National Agencies in Trade Related Capacity Build ing." *World Economy* 25: 1235 61. Links trade

related capacity building to countries' trade and de
velopment strategies.

de Velde, Dirk Willem, Massimiliano Calì, Adrian Hewitt,
and Sheila Page. 2006. "A Critical Assessment of the
EU's Trade Related Technical Assistance to Third
Countries   Lessons from the Past, Policy Options for
the Future." London: Overseas Development Institute.
An outside assessment of trade related capacity building
efforts of the European Union.

World Bank, Independent Evaluation Group. 2006. "As
sessing World Bank Support for Trade, 1987 2004."
Washington, DC: World Bank. An outside assessment
of trade related capacity building efforts of the World
Bank. Available at http://www.worldbank.org/ieg/trade.

World Trade Organization. 2005. "Strategic Review of
WTO provided TRTA Activities." Geneva: WTO. A
critical assessment of technical assistance activities pro
vided by the WTO Secretariat.

MICHEL KOSTECKI

■ trade-related intellectual property rights (TRIPS)

See Agreement on Trade Related Aspects of Intellectual
Property Rights (TRIPS)

■ trade-related investment measures (TRIMs)

Governments adopt measures to attract foreign in
vestment and to encourage the use of foreign in
vestment in accordance with national priorities. Such
measures that affect trade are known as *trade-related
investment measures*. TRIMs designed to attract in
vestment include fiscal incentives, tax rebates, and
preferential terms on land and service. Common
TRIMs designed to encourage investment in accor
dance with national priorities include local content
requirements (requiring investors to use local inputs
in production) and export performance require
ments (requiring investors to export a certain pro
portion of their output). TRIMs also include trade-
balancing requirements, foreign exchange balancing
requirements, exchange restrictions, domestic sales
requirements, manufacturing requirements, product
mandating requirements, manufacturing limita-
tions, technology transfer requirements, licensing
requirements, remittance restrictions, local equity
requirements, and employment restrictions.

The primary reason countries adopt TRIMs is to
promote development. For example, some countries
adopt TRIMs to foster domestic industries (through
local content requirements), to expand exports
(through export performance requirements), and to
remedy deteriorating balance of payments (through
foreign exchange balancing requirements). Coun
tries also adopt TRIMs to address restrictive business
practices and anticompetitive behavior of multina-
tionals. One means is through technology transfer
policies. For example, licensing requirements stipu-
late that foreign investors license technologies to host
firms; and technology transfer requirements oblige
foreign investors to transfer technologies on non-
commercial terms or to conduct research and de-
velopment locally.

The evolution of TRIMs as a policy tool coincides
with changes in trade and investment. Since the late
1980s, foreign direct investment (FDI) has grown
worldwide. According to the United Nations Con
ference on Trade and Development (UNCTAD),
FDI increased from $203 billion in 1990 to $735
billion in 2001. Developing countries received $238
billion of this investment in 2001. Accordingly de
veloped countries are the predominant sources and
recipients of FDI. Further, FDI tends to be *horizontal*
in that the output of foreign affiliates is sold in the
foreign country. As FDI has grown, trade within
companies has also grown. This trade (referred to as
intrafirm trade) includes trade between subsidiaries
located in different countries and between a subsid-
iary and its headquarters. According to the World
Trade Organization (WTO), approximately one-
third of the $6.1 trillion total of world trade in 1995
was intrafirm trade. Consequently since the 1980s an
increasingly large share of trade is *related* to investment.

History of Agreements and Treaties Before
1995 few international frameworks provided disci-
plines for foreign investment. During the Uruguay

Round (1986 95) of the General Agreement on Tariffs and Trade (GATT), the United States proposed negotiations on policies that affect FDI. Weak support for these proposals, particularly among developing countries, resulted in negotiations on the narrower concept of trade-related investment measures.

At the conclusion of the Uruguay Round in 1995, members adopted the Agreement on Trade-Related Investment Measures (TRIMs Agreement) as one of four agreements of the WTO treaty. The agreement requires countries to phase out policies identified as inconsistent with GATT rules on national treatment and quantitative restrictions. TRIMs identified as inconsistent with national treatment rules include local content requirements and trade-balancing requirements. TRIMs identified as inconsistent with rules on quantitative restrictions include trade-balancing requirements constituting restrictions on imports, exchange restrictions resulting in restrictions on imports, and domestic sales requirements involving restrictions on exports.

On adoption in 1995, the TRIMs Agreement required member states to notify the WTO of existing TRIMs within 90 days. Members were then given a transition period to phase out the notified TRIMs (two years for developed countries, five years for developing countries, and seven years for least-developed countries). The agreement allows exceptions for developing and least-developed countries and permits the extension of the transition periods. For countries acceding to the WTO since 1995, obligations depend on their accession terms.

In addition to the TRIMs Agreement, other WTO agreements address specific aspects of TRIMs. These include the agreements on Trade-Related Aspects of Intellectual Property Rights, on Subsidies and Government Procurement, and the General Agreement on Trade in Services. However no comprehensive framework on TRIMs exists within the WTO to bridge these agreements.

Given its limited scope, the TRIMs Agreement was to be reviewed within five years to consider additional provisions. The 1996 WTO Ministerial Conference in Singapore initiated this process by calling for discussions in four areas (the "Singapore issues") including trade and investment, and by establishing working groups to facilitate analyses. The 2001 Ministerial Conference in Doha then produced the Doha Development Agenda (DDA), which mandated negotiations on new rules to begin after the 2003 Conference pursuant to a consensus on modes of negotiations. However the 2003 Ministerial Conference in Cancun ended without a consensus on several Singapore issues. In 2004 members dropped trade and investment from the DDA. Other efforts to create a multilateral agreement on investment include the Organisation for Economic Cooperation and Development's Multilateral Agreement on Investment. Negotiations on this agreement began in 1995, but subsequently stalled when participants were unable to reach a compromise.

The greatest challenge to creating a comprehensive framework is the divergent interests of countries. Those in favor of a new agreement on trade and investment (primarily developed countries) argue that the existing regime of bilateral and regional arrangements is confusing. They argue that a multilateral agreement would create a stable nondiscriminatory environment that would increase investment. Those against (primarily developing countries) argue that the existing regime provides adequate protections for multinationals and question whether a multilateral agreement would increase investment. Critics claim that a new multilateral agreement would add obligations while limiting countries' ability to align investment with development objectives.

See also domestic content requirements; foreign equity restrictions; international investment agreements

FURTHER READING

Balasubramanyam, V. N. 1991. "Putting TRIMs to Good Use." *World Development* 19 (9): 1215 24. Examines this relationship and argues that some (but not all) TRIMs promote development objectives.

Chao, Chi Chur, and Eden S. H. Yu. 2000. "TRIMs, Environmental Taxes, and Foreign Investment." *Canadian Journal of Economics* 33 (3): 799 817. Considers the welfare effects of combinations of TRIMs and trade and environmental policies.

. 2003. "Export Performance Requirements, For eign Investment Quotas, and Welfare in a Small Dy namic Economy." *Journal of Development Economics* 72 (1): 387 400. Examine the welfare effects of specific TRIMs, including foreign investment quotas and ex port share requirements.

Correa, Carlos, and Nagesh Kumar. 2004. *Establishing International Rules for Foreign Investment: Trade Re lated Investment Measures (TRIMS) and Developing Countries*. London: Zed Books. Considers the per spective of developing countries in a variety of in itiatives to establish and reform international rules for foreign investment, including a new multilateral frame work within the WTO.

Greenaway, David. 1992. "Trade Related Investment Mea sures and Development Strategy." *Kyklos* 45 (2): 139 59. Argues that TRIMs are second best policies because they emerge as a consequence of preexisting distortions.

Greenaway, David, and Andre Sapir. 1992. "New Issues in the Uruguay Round: Services, TRIMs and TRIPs." *European Economic Review* 36 (2 3): 509 18. Discusses the motivation for including TRIMs in the Uruguay Round and whether international rules should be ex tended to embrace TRIMs.

PAMELA J. SMITH

■ transaction costs

See foreign market entry

■ transfer pricing

Transfer pricing decisions arise when one division of a firm sells goods or services to another autonomous division within the same firm. The firm's best interest is served when it selects a transfer price that maximizes total firm profits, which are generated by the aggregate efforts of the firm's separate divisions. Selecting the ideal transfer price, however, requires the firm to consider many facets and implications of its decision, including costs, incentives, and the details of the policy environment.

Firms often organize themselves as multidivision entities to reap the benefits of decentralized decision making. For example, in addition to the firm's headquarters, which coordinates the firm's activities and sets the rules for its subsidiaries, the firm might create separate divisions for parts production, assembly, research and development, advertising, and distribution. Thus transfer prices must be set when components from the parts division are provided to the assembly division, or when the final goods from the assembly division are supplied to the firm's distribution arm.

Hirshleifer's (1956) pioneering work on transfer pricing demonstrated that when there is no separate market for intermediate inputs, the transfer price should be set to the marginal cost of the producing division. In contrast, if there is a perfectly competitive market for intermediate inputs, the transfer price should be set at the market price for the intermediates. Aside from these simple cases, however, the optimal transfer price depends on a number of economic fundamentals, including the nature of competition in intermediate and final good markets, the structure of firm costs, and whether there is demand or technical dependence between separate divisions of the firm. As a result, firms not only must ask what the optimum transfer price is, but also, as noted by Holmstrom and Tirole (1991), need to determine whether it is best to organize the firm vertically, placing the coordination of all price decisions at the top.

When a firm locates its divisions in different jurisdictions either different countries or different fiscal entities (states, provinces, or prefectures) within a country differences in jurisdictional tax rates may influence transfer prices; firms have an incentive to maximize firm profits by strategically manipulating transfer prices as a means of locating profits in low-tax countries or locations. For example, consider a firm that produces inputs in a low-tax country that are used for assembly in a high-tax country. This firm can reduce its tax payments if it increases the declared value of the parts it sells to its assembly division in the high-tax country, because the price manipulation increases revenue and profits in the low-tax country while it increases costs and reduces profits in the high-tax assembly country. Similarly if the country tax environment is reversed such that the corporate tax rate is higher in the country where parts are produced

and lower in the country where the firm does its assembly, the firm could reduce its tax burden by reducing the transfer price it declares on the parts it exports from the parts affiliate to the assembler in the low-tax country.

A full understanding of the incentives to manipulate transfer prices requires more information yet on the organization of the multinational and the firm's environment. For example, the simple case of transfer price manipulation suggests that a firm will reduce its current tax payment by the change in the declared value of its exports multiplied by the tax differential between the sending and receiving countries. Depending on the form of taxation in the firm's headquarters country, however, remaining taxes not paid in the current period may come due when income from a firm's overseas subsidiaries is repatriated to the firm's headquarters. Further, if the definition of taxable income differs across countries, the firm will have to evaluate the effect of its changed declaration on its worldwide tax payments. Finally, a firm's decision to engage in tax-driven transfer price manipulation may trigger other policy considerations. Along these lines, Horst (1971) describes how multinationals need to consider the effect of ad valorem tariffs, which provide a disincentive to increasing declared export prices and an incentive to underreporting of the export price. Further, when a horizontal multinational sets its transfer price, it may face further constraints on its choice of prices, since arbitrage conditions define the maximum degree of cross-country price differences that are sustainable.

Since tax authorities are aware that multinational firms may have strong financial incentives to manipulate their declared transfer prices to shift income from highly to more lightly taxed countries, most governments provide guidance on the setting of transfer prices for international transactions. For example, U.S. transfer price regulations instruct firms to use transfer prices for their internal transactions that are the same as the prices that would apply if the firm were conducting the sale at arm's length, as is the case when the firm sells the part or service to an unrelated party. Thus the ability of multinational firms to engage in tax-induced transfer price manipulation is limited when firms sell homogeneous goods or services. When firms do not have arm's length transactions to guide them in their choice of transfer prices, they are advised to use cost-plus or comparable profits as alternative methods for setting transfer prices. Although these guidelines prevent firms from declaring any arbitrary price, they nonetheless provide some latitude in price setting; the fact that many accounting firms have large divisions dedicated to transfer pricing attests to the complexity of these decisions, as well as the value associated with these choices.

Although transfer pricing decisions can occur in either a purely domestic or an international context, the nature of international investment makes transfer pricing a particularly salient issue. For example, due to international differences in comparative advantage, it often is attractive for firms to place their assembly facilities in low-wage, labor-abundant countries while they conduct more capital-intensive activities at home or in a capital-abundant country. Alternatively, in the case of horizontal investment, firms may decide to set up overseas production sites as a means of reducing variable transportation and tariff costs. In either case, firms that operate as multinationals are generally more productive than firms that are purely domestic, and it is the possession of intangible assets   designs, blueprints, trademarks, proprietary management systems   that enables firms that go multinational to expand across national borders successfully. As a result, the fact that intangible assets are generally present in and responsible for multinational firm activity implies that multinational firms will have more scope to manipulate transfer prices. At the same time, transfer pricing is not the only avenue for tax minimization, as multinational firms can reduce their worldwide payments through other means, such as their allocation of debt, timing of dividends, or   in the case of compensation for transferred intangible assets   use of royalty payments or licensing fees.

Finally, although there are many avenues for tax minimization by multinational firms, transfer price manipulation appears to be one that is actively pursued. Grubert and Mutti (1991) provide empirical

evidence of this from U.S. multinational firms, showing that variation in the location of reported profits is consistent with income-shifting incentives and that real investment activity and firm exports of multinational firms are influenced by international differences in corporate taxation. Similarly, Swenson (2001) and Clausing (2003) find evidence in U.S. trade data that variation in trade prices is consistent with transfer pricing motives.

See also intangible assets; multinational enterprises

FURTHER READING

Clausing, Kimberly A. 2003. "Tax motivated Transfer Pricing and U.S. Intrafirm Trade Prices." *Journal of Public Economics* 87 (9): 2207 23. Empirical evidence that transfer prices respond to incentives.

Grubert, Harry, and John Mutti. 1991. "Taxes, Tariffs, and Transfer Pricing in Multinational Corporate Decision Making." *Review of Economics and Statistics* 73 (2): 285 93. Documents that multinational firm decisions are consistent with tax minimization incentives.

Hirshleifer, Jack. 1956. "On the Economics of Transfer Pricing." *The Journal of Business* 29 (3): 172 84. Pio neering paper examining optimal transfer price decisions for firms.

Holmstrom, Bengt, and Jean Tirole. 1991. "Transfer Pric ing and Organizational Form." *Journal of Law, Eco nomics, and Organization* 7 (2): 201 28. Places transfer pricing decision in its larger context, which includes organizational form.

Horst, Thomas. 1971. "The Theory of the Multinational Firm: Optimal Behavior under Different Tariff and Tax Rates." *Journal of Political Economy* 79 (5): 1059 72. Theoretical analysis of optimal transfer price setting by firms.

Rugman, A. M., and Lorraine Eden, eds. 1985. *Multi nationals and Transfer Pricing*. New York: St. Martin's Press. Thorough reference on issues related to firm transfer pricing decisions.

Swenson, Deborah L. 2001. "Tax Reforms and Evidence of Transfer Pricing." *National Tax Journal* 54 (1): 7 25. Provides empirical evidence that reported transfer prices appear to reflect incentives based on tariffs and taxes.

DEBORAH L. SWENSON

■ **transfer problem**

International financial transfers take many forms. The classic example is war reparations, such as the payments by France after the 1870 Franco-German War and by Germany after World War I. The same general principles apply to other types of interna tional transfer. These include foreign aid to devel oping countries, capital transfers from the richer to the poorer members of the European Union, and the remittances sent home by emigrants. Other pertinent cases include commodity producers that enjoy a windfall improvement in export prices (think of the impact of an increase in world oil prices on oil pro ducers) and countries with net external liabilities that must make interest payments to foreign creditors.

The transfer problem relates to the impact of an international transfer on the structure of exchange rates and relative prices. To evaluate the full impact on the economies of the countries that pay and re ceive the transfer we need to assess the transfer problem.

There are circumstances under which an inter national transfer has no price impact. In the classic formulation of Keynes (1929a), "If £1 is taken from you and given to me and I choose to increase my consumption of precisely the same goods as those of which you are compelled to diminish yours, there is no Transfer Problem." It is implicit in this formu lation, however, that the wealth transfer has no im pact on supply capacities in either economy. This is open to question, since there is a natural tendency for an increase in wealth to be associated with a reduction in aggregate work effort. Accordingly, we may expect a decline in the level of production of the recipient economy relative to the level of production in the sender economy. To the extent that the two econo mies specialize in different goods, this shift in relative production levels may be expected to raise the prices of the goods produced by the recipient economy relative to the prices of the goods produced by the sending economy, even if consumption patterns are identical in both economies.

Moreover it is unrealistic to assume that house holds in both economies have the same consump tion patterns. First, transportation costs means that

domestically produced goods sell for lower prices to domestic consumers than to foreign consumers, which naturally tilts domestic spending patterns toward home-produced goods. For this reason, a transfer of wealth from the foreign country to the domestic country will be associated with an increase in the demand for home-produced goods and a decline in the demand for foreign-produced goods, leading to an increase in the relative price of home goods in terms of foreign goods. From the perspective of the home country, this implies an improvement in the terms of trade the price obtained for exports improves relative to the price of imports.

Second, all countries produce a wide range of goods and services that are sold exclusively to domestic residents. A transfer of wealth from the foreign country to the home country means that demand for such nontraded goods rises at home and declines overseas. If the supply of such goods is somewhat inelastic, this will translate into an increase in the relative price of nontradables in the home country and a decline in the relative price of nontradables in the foreign country.

Third, there is considerable evidence that firms in some industries are able to set different prices in different markets, through segmentation strategies that make it difficult to arbitrage price gaps. An intuitive pattern is for firms to charge higher prices in wealthier markets. It follows that a transfer that enriches the home country relative to the foreign country may result in an increase in the prices of such goods at home relative to the foreign market. Each of these factors points to an increase in the domestic price level relative to the overseas price level the recipient country experiences real exchange rate appreciation (equivalently, the sender country experiences real exchange rate depreciation).

The scale of the real exchange rate movements can be mitigated by several factors, however. For instance, international capital flows and the intersectoral mobility of labor between the traded and nontraded sectors may eliminate the impact of demand factors on the relative price of nontraded goods: if there is an increase in demand for nontraded goods, workers and entrepreneurs can switch from the export sector to the nontraded sector in order to meet the expansion in demand and capital can flow in from overseas to provide extra capacity. This supply response is more likely to apply over the medium term, with factors unlikely to move across sectors quickly or in response to purely temporary surges in demand. Similarly the supply response will be stronger for transfers that are signaled in advance, rather than unanticipated windfalls. Some factors (such as land) have limited supply elasticity, however, such that it remains likely that the level of demand influences the relative price of nontradables even over the long term.

In addition, an increase in demand for domestic tradables may in part be met by the entry of new firms that produce new varieties of goods this is known as the "extensive" margin of adjustment. In this case, the expansion in demand has a smaller impact on the conventionally measured terms of trade, since the "extensive" margin of trade expands. This is most likely in advanced economies in which an active research and development sector fosters innovation.

Moreover, if countries have the ability to borrow and lend in international financial markets, the recipient may opt to save much of a temporary transfer by accumulating foreign assets. This strategy would allow the recipient to maintain a permanently higher level of consumption even after the expiration of the transfer, since the investment income on the accumulated assets would still be available. For this reason, countries such as Norway and Kuwait have opted to park much of their oil revenues in national investment funds, rather than enjoy a temporary boom in consumption that could not be sustained. A further advantage from this strategy is that it may help a recipient avoid the "Dutch Disease" syndrome whereby the increase in domestic wage costs that is precipitated by a surge in domestic demand leads to a contraction in the domestic export sector, which is costly if that sector has the greatest potential for delivering long-term productivity growth.

Unfortunately many developing countries have experienced problems in managing resource inflows (whether windfalls from surges in commodity prices

or foreign aid receipts), such that there is a domestic demand surge and a sharp initial real appreciation. This is compounded by a propensity to increase external borrowing on the back of an increase in wealth, with the political pressures to increase spending exceeding the scale of the transfer. This "voracity effect" means that such countries experience macroeconomic instability, with the initial boom followed by a period of debt repayment, depressed consumption, and a decline in the real exchange rate (Tornell and Lane 1999).

Finally, we should note that the scale of the transfer problem may be expected to vary across economies of different sizes. In particular, the nontraded sector is relatively bigger in economies with larger populations and those that are more geographically isolated; larger economies may also have greater market power in international markets, such that their terms of trade respond more strongly to shifts in demand. For this reason, the empirical evidence shows that the transfer problem is much less important for small, open economies than for continental-sized economies (Lane and Milesi-Ferretti 2004). Even for small, open economies, however, the transfer problem remains central to the analysis of international resource flows.

See also capital mobility; nontraded goods; real exchange rate; terms of trade

FURTHER READING

Keynes, John Maynard. 1929a. "The German Transfer Problem." *Economic Journal* 39: 1 7. Keynes proposes that the payment of war reparations by Germany to France necessarily involves a reduction in the German price level relative to the French price level.

———. 1929b. "A Rejoinder." *Economic Journal* 39: 179 82. Keynes clarifies his argument, in response to the critique provided by Bertil Ohlin.

Lane, Philip R., and Gian Maria Milesi Ferretti. 2004. "The Transfer Problem Revisited: Real Exchange Rates and Net Foreign Assets." *Review of Economics and Statistics* 86: 841 57. Provides empirical evidence in support of the transfer effect for a large group of countries.

Ohlin, Bertil. 1929. "Transfer Difficulties: Real and Imagined." *Economic Journal* 39: 172 78. Ohlin argues that there are scenarios under which an international transfer need not affect relative price levels.

Tornell, Aaron, and Philip R. Lane. 1999. "The Voracity Effect." *American Economic Review* 89: 22 46. Analyzes the voracity effect, by which a divided society increases spending more than proportionally in response to a transfer.

PHILIP R. LANE

■ transition economies

Transition economies are those that are in the process of moving from central planning to markets. A centrally planned economy, sometimes also known as a command economy, is one where the means of production are publicly owned, the division of total product between investment and consumption is a centrally made political decision, and all economic activity is controlled in accordance with a central plan that assigns quantitative production goals and allots raw materials to enterprises. A market economy, on the other hand, is one where the means of production are mostly privately owned, the production and consumption of goods result from decentralized decisions made by individuals in pursuit of their own ends in light of prevailing prices, and prices adjust to equate supply and demand.

These definitions describe idealized types; in practice, most economies incorporate elements drawn from both, with the mixture determining whether they are closer to centrally planned or market economies. That said, an event of seminal significance in the late 20th century was the transition from plan to market in the countries of Eastern Europe and the former Soviet Union. An economic transition has also been underway in China and Vietnam. The former group of countries has, however, mostly seen transition as a broader concept, embodying a political transition from communism to multiparty democracy alongside the economic transition. This entry focuses on the economic transition in Eastern Europe and the Commonwealth of Independent States. Eastern Europe comprises Albania, Bosnia and Herzegovina, Bulgaria, Croatia,

the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Macedonia, Montenegro, Poland, Romania, Serbia, Slovakia, and Slovenia. The Commonwealth of Independent States (CIS) includes Armenia, Azerbaijan, Belarus, Georgia, Kazakhstan, the Kyrgyz Republic, Moldova, Russia, Tajikistan, Turkmenistan, Ukraine, and Uzbekistan.

An Analytical Framework The fall of the Berlin wall in 1989 heralded the demise of communism in Eastern Europe, whereas the exit from communism did not occur until 1991 in the former Soviet Union. A key first element signaling economic transition was the liberalization of most prices, so that households and firms henceforth faced international prices for tradable goods. However, large parts of the planned economy transacted at subsidized prices based on cheap energy and subsidized transport. Price liberalization, together with the disruption of the organizational arrangements governing production and trade under central planning, implied that many enterprises inherited by the transition countries were no longer viable. This found expression in what has come to be known as the "transition recession," during which gross domestic product (GDP) is estimated to have fallen by about 15 percent in Eastern Europe and by 40 percent in the CIS, reflecting in part the more distorted structure of the latter and, in the case of some of the poorer ex-Soviet republics, the cessation of budget transfers from Moscow. These numbers are not uncontroversial because the GDP data for the early years, by focusing on large enterprises where output had been declining and not adequately incorporating new businesses and an emerging informal sector, likely overestimate the severity of the transition recession. But this does not modify the qualitative thrust of the observations made here.

An important second element consisted of policy and institutional measures to elicit a supply response that would allow output to recover. These included (1) limiting state assistance (such as tax exemptions, fiscal and financial subsidies, budget and tax offsets, and directed credits) to enterprises, thereby promoting their restructuring or closure and making possible the transfer of assets and labor to viable enterprises; (2) preventing the theft of state assets, through either (a) privatization and an enforceable set of property rights in a private ownership economy, or (b) administrative methods in the few countries where centralized political structures had remained substantially intact; and (3) creating a business environment that did not discriminate among firms, whether old, restructured, or new, and comprised elements such as business licensing and registration, tax policy and administration, and a legal-cum-judicial system protective of property rights.

While restructuring and reallocation of resources in the enterprise sector constitutes the essence of the transition, implementation of this agenda required attention to two other sets of policies. Price liberalization implied that inflation, hitherto repressed in centrally planned economies through unavailability of consumer goods, and expressed in the form of forced monetary savings, had the potential to erupt into hyperinflation. This called for a third element in the set of policies, macroeconomic stabilization, to bring inflation under control.

Fourth, since restructuring and reallocation of resources involved the loss of workers' livelihood, social safety nets needed to be used   fiscal resources permitting   to finance retraining, early retirement, and cash assistance to bridge the period between job destruction and job creation and thus lower the political costs of the transition. This quartet of measures   (1) price liberalization; (2) creation of conditions for the growth of the private sector, including hard budget constraints on enterprises, a nondiscriminatory investment climate, and institutions to control managerial behavior; (3) macroeconomic stabilization; and (4) creation of social safety nets   made up the core of economic policy on managing the transition in Eastern Europe and the former Soviet Union from central planning to a market economy during the early years.

Initial Conditions or Policy Reforms? GDP in 2006 compared with 1990 was 50 percent higher in Eastern Europe and had nearly caught up with its 1990 level in the CIS. This was an increase of some three-quarters from the nadir of the transition re-

cession in each subgroup of countries. An extensive debate surrounds the identification of the causes of the transition recession and the subsequent recovery from it. Cross-country statistical analysis suggests that initial conditions such as the degree of industrialization, trade dependence on other communist countries, repressed inflation, and black market exchange rates during the communist period were closely associated with lower growth during the early years of the transition. On the other hand, policy reforms, including elimination of central planning and mandated allocations through government orders, macroeconomic stabilization, trade liberalization, and competition, as well as a history of prior experience with market reforms and nationhood, were closely associated with subsequent growth performance, becoming stronger as the transition progressed.

The positive association between reforms and growth may seem surprising when the period of analysis includes recent years because the CIS, which has generally made less progress with reforms, has been growing particularly rapidly, more so than Eastern Europe since the Russian financial crisis of 1998. However, the positive impact of reforms on growth indeed continues to hold once two factors are controlled for: first, that the rapid growth in the CIS reflects in part catch-up from a deep recession, and second, that a surge in the price of oil has benefited the CIS oil exporters and other CIS countries through their trade linkages with those exporters (European Bank for Reconstruction and Development 2004).

Country Patterns of Transition The patterns emerging from cross-country statistical analysis need to be complemented by examining the experience of individual countries or subgroups of countries. The transition countries are highly diverse, with per capita incomes in 1991 ranging from an estimated $96 in Armenia to $6,368 in Slovenia. Although countries pursued different combinations of policies in the early years of transition, some broad generalizations are possible regarding the quartet of measures previously identified as key to managing the transition. Macroeconomic stabilization packages including,

among other things, tight monetary and credit policies, wage control policies, budget deficits financed from noninflationary sources, and suitable choice of exchange rate regimes, had been put in place in virtually all the transition countries by 1995. This helped bring down inflation, which had ranged from 26 percent in Hungary to a hyperinflationary 57,000 percent in Georgia prior to stabilization, to single digits in most countries by 1998 (Fischer and Sahay 2000).

Turning to enterprise sector reform, Estonia, Hungary, and Poland exemplify the imposition of hard budget constraints and the creation of an investment climate conducive to the entry of new firms in the early years of transition. This was less evident in those years in the Czech and Slovak republics and Lithuania, and it was only later that harder budget constraints were introduced, whereupon faster restructuring helped set these economies on the path followed by the earlier set of countries. In contrast, Bulgaria, Moldova, Romania, Russia, and Ukraine liberalized their economies but failed to impose hard budget constraints or contain the theft of assets either through law or administrative control. Russia and Ukraine encouraged new entry early in the transition, but the capture of the state by narrow vested business interests, comprising both old enterprises and well-connected early entrants, created a poor investment climate and discouraged further entry by potential competitors. Countries such as Belarus, Uzbekistan, and Turkmenistan, which retain strong elements of central planning, neither liberalized prices nor hardened budget constraints, but the survival of a highly centralized political structure allowed the apparatus of the command economy to be used to limit the extent of asset stripping by enterprise managers that had proved so damaging to growth elsewhere in the CIS.

Finally, adequate fiscal resources allowed countries in Central Europe to use social safety nets to cushion the impact of transition on displaced workers through retraining or early retirement. This option was not available to countries in the CIS due to a major loss of revenue from those state enterprises that were no longer profitable in a market economy.

The lack of a functioning social safety net implied that the deindustrialization that occurred in transition economies, which were typically overindustrialized compared with market economies at similar income levels, led to a movement of workers away from industry and into subsistence agriculture.

Institution Building The far-reaching nature of changes characteristic of the transition from a centrally planned to a market economy, particularly illustrated by the challenges of reforming the enterprise sector, implied that the countries of Eastern Europe and the former Soviet Union faced a formidable agenda of institutional reform. While this fact was recognized early, progress in creating the institutions of a market economy has lagged those reforms that could be accomplished by stroke-of-the-pen policy changes in all countries. As an example, this distinction is captured for the private and financial sectors by the European Bank for Reconstruction and Development's annual reports on transition. The EBRD defines "initial phase" reforms to include small-scale privatization, price liberalization, and trade and foreign exchange liberalization. "Second phase" reforms are defined to include large-scale privatization, governance and enterprise restructuring, competition policy, financial sector development, and infrastructure, the implementation of which calls for deep institutional reform. Initial phase reforms are virtually complete in the countries of Central Europe and the Baltic states that joined the European Union in 2004 and, to a lesser extent, in Bulgaria and Romania, which became European Union members in 2007. In contrast, second phase reforms lag virtually everywhere and especially in Eastern Europe outside the European Union and the CIS countries. In the context of economic integration, for example, most transition countries have generally liberal trade policy regimes. However, the countries where second phase reforms are lagging have not been able to attract the foreign direct investment that would enable deeper integration into global production and distribution chains and improve their longer-term prospects for economic growth. This is due to insufficient progress in developing institutions capable of implementing competition policy; effective regulatory supervision of services such as banking, telecommunications, and transport; as well as trade facilitation (World Bank 2005b).

Developing the Private Sector The share of the private sector in GDP, which is estimated to range from 25 percent in Belarus to 80 percent in Estonia, the Czech Republic, and Hungary, is often regarded as an important indicator of progress in transition. The private sector reflects both the results of privatization as well as entry and exit of privately owned firms that had never been owned by the state. Together with price liberalization and macroeconomic stabilization, privatization has been an important and widely discussed element in the transition from a command to a market economy. Privatization in the early years of transition was seen as a way of generating incentives for the new owners to use enterprise assets efficiently. This consideration was particularly important in the CIS countries, where weakening governments were unable to prevent asset stripping by enterprise managers, which had begun well before the exit from communism in the former Soviet Union. And, politically, privatization also helped ensure the irreversibility of the transition in its early years by creating a class of owners with a stake in the market economy.

While the privatization of small-scale enterprises was relatively uncontroversial, policymakers were faced with a difficult choice regarding the management of assets in medium- and large-scale enterprises. On the one hand, privatization in the absence of institutions of corporate governance could not guarantee that assets would not be stolen or that the rights of minority shareholders would be protected by the new owners. On the other hand, as noted, neither was the success of continued state ownership in an environment of weak government control assured. The problem of adequate monitoring of managerial behavior by shareholders could be overcome to some extent in cases where enterprises could be sold to concentrated private owners, that is, a few owners with larger stakes rather than many stakeholders with smaller stakes in the enterprise. However, the type of concentrated ownership mattered as

well. Enterprises sold to strategic investors equity capital investors that usually operate in the same or similar field and have a long-term interest in developing the company particularly if they were foreign investors, performed much better than those controlled by holding companies or other financial institutions. The selection of strategic investors mattered too. Enterprises sold through transparent tenders or auctions generally attracted better owners, outperforming enterprises sold directly to politically connected parties, frequently at highly subsidized prices. Without such safeguards, concentrated ownership could not avoid the risk of expropriation of assets and income belonging to minority shareholders.

Many of the dilemmas surrounding medium- and large-scale privatization arose when the preferred method of privatization direct sales through transparent tenders or auctions to strategic investors was unavailable and a minimum of institutional capacity to prevent asset stripping by enterprise managers in the interim period was absent. Indeed the evidence suggests that privatization to concentrated outside owners has benefited restructuring whereas privatization to diffuse owners and to enterprise owners and managers has not (Djankov and Murrell 2002). Reforming laws and institutions to protect investors and monitor managerial behavior in countries with no market experience was rendered even more difficult when opposed by early winners from transition, who did so because such reforms would dissipate the rents accruing to them. In Russia, for example, powerful insiders frequently hampered enforcement efforts by the Securities and Exchange Commission. Hence in an environment of substantial capture of the state by narrow private interests, privatization did not create enough demand for the enforcement of property rights, contrary to early expectations. Indeed, while privatization has been found to be positively associated with a measure of public governance, namely, the state's capacity to provide key public goods in low-state-capture environments, the association turns out to be negative in high-state-capture environments (European Bank for Reconstruction and Development 1999).

The development of the private sector is also shaped by firm turnover entry and exit which has been substantial in the transition economies. The role of hard budget constraints on all enterprises and a business environment that does not discriminate between firms has been of particular importance in this regard. This is because the entry of new firms, which were able to occupy market niches that did not exist under central planning, as well as the exit of obsolete firms that were adapted to the command economy, contributed relatively more to productivity growth in transition economies compared with industrial and other developing countries. Furthermore, hard budget constraints and a nondiscriminatory business environment were mutually reinforcing. This is because hard budget constraints led to job destruction, which would have been politically difficult to withstand in the absence of job creation by new firms, which required, inter alia, a nondiscriminatory business environment. Early reformers such as Estonia and Hungary provide examples of job creation that was able to offset job destruction to a greater extent than in late reformers. Conversely the continued subsidization of unprofitable state-owned firms (soft budget constraints) could result in macroeconomic crises, as was the case in Bulgaria and Romania in the mid-1990s, resulting in a business environment that was not conducive to the development of new firms. Hence it is important that the share of the private sector in GDP as an indicator of progress in transition be complemented by an assessment of the extent to which the business environment provides incentives to privatized and new private firms to add value rather than engage in theft and rent seeking.

The Political Economy of Transition Why were most countries in Central Europe and the Baltic states able to implement more comprehensive reforms than those in the CIS? The answer depends, in part, on whether a government could credibly commit to a comprehensive reform program. This commitment would give potential winners such as owners of entrant firms and workers the confidence that they would eventually realize high enough returns to offset the initial adjustment costs incurred by

them in exiting the state sector and entering the private sector. As reforms progressed, the expanding private sector created gains that were large enough to offset the losses of both (1) workers in a downsizing state sector and (2) early winners from partial reforms who had already established positions of advantage and whose gains would be competed away in the event of more comprehensive reform.

In former Soviet states such as Russia and Ukraine, the collapse of communism was rooted in a contest among competing elites rather than in any broad social movement. Governments, which had been captured by early winners from activities such as asset stripping and arbitrage between different sets of prices during price liberalization, lacked the credibility to build and sustain political support for a comprehensive reform program. This discouraged entry by competing groups that did not believe that the reform process would proceed far enough to deliver them a return net of adjustment costs. In much of Central Europe and the Baltic states, by contrast, political institutions emerged from round-table negotiations among broadly representative popular fronts in the aftermath of revolutions against communist rule. This, together with the pull of potential European accession that has been a powerful driver of political and economic institutional development, contributed to a wider social consensus on the main directions of reform and broad public support for comprehensive reform programs in the early stages of transition. These developments allowed those governments to be seen as credibly committed to such programs. As constituencies with a stake in reforms grew stronger, those groups that had gained an early advantage from liberalization and privatization were unable to convert those gains to erect barriers to competition and entry (World Bank 2002).

See also economic development; international income convergence; international institutional transfer; political economy of policy reform; Washington consensus

FURTHER READING

de Melo, Martha, Cevdet Denizer, and Alan Gelb. 1996. *From Plan to Market: Patterns of Transition*. Washington, DC: World Bank.

de Melo, Martha, Cevdet Denizer, Alan Gelb, and Stoyan Tenev. 1997. *Circumstance and Choice: The Role of Initial Conditions and Policies in Transition Economies*. Washington, DC: World Bank. The two papers by de Melo et al. introduced the Liberalization Index, a widely used measure of the extent of policy reform, and show that the extent of liberalization is an important determinant of economic growth during the transition.

Djankov, Simeon, and Peter Murrell. 2002. "Enterprise Restructuring in Transition: A Quantitative Survey." *Journal of Economic Literature* 40 (3): 739 92. Surveys the empirical literature analyzing the process of enterprise restructuring in transition economies.

European Bank for Reconstruction and Development. 1999. *Transition Report*. London: EBRD. Takes stock of the first ten years since the fall of the Berlin wall and examines the forces that moved transition forward as well as the constraints that held back progress.

———. 2004. *Transition Report*. London: EBRD. Reviews the relationship between reforms and growth in transition countries, including the role of factors such as the depth of the transition recession and rising oil prices.

Fischer, Stanley, and Ratna Sahay. 2000. *The Transition Economies after Ten Years*. IMF Working Paper No. 00/30. Washington, DC: IMF. Provides a summary of the macroeconomic performance of the transition economies.

World Bank. 1996. *World Development Report: From Plan to Market*. Washington, DC: World Bank. This report, written during the middle of the transition's first decade, assesses (1) how countries grappled with the initial dilemmas of transition and (2) how countries could best consolidate initial reforms to create institutions supportive of a market economy.

World Bank. 2002. *Transition: The First Ten Years Analysis and Lessons from Eastern Europe and the Former Soviet Union*. Washington, DC: World Bank. Reviews the roles of policy, political economy, and external shocks in transition economy growth performance.

World Bank. 2005a. *Enhancing Job Opportunities: Eastern Europe and the Former Soviet Union*. Washington, DC: World Bank. Addresses why labor market outcomes have been disappointing during the transition and suggests policy interventions that can foster job creation and reduce unemployment.

World Bank. 2005b. *From Disintegration to Reintegration: Eastern Europe and the Former Soviet Union in Interna tional Trade*. Washington, DC: World Bank. Analyzes how Eastern Europe and the former Soviet Union have been reintegrated into international commerce since the disintegration of the Soviet trade bloc.

World Bank. 2008. *Unleashing Prosperity: Productivity Growth in Eastern Europe and the Former Soviet Union*. Washington, DC: World Bank. Examines the micro foundations of economic growth since 1998 and estab lishes that policy reforms that promote governance and microeconomic stability, market competition, infra structure quality, financial deepening, labor market flex ibility, and skill upgrading are important for achieving higher productivity growth.

PRADEEP K. MITRA

■ **transnational corporations**

See multinational enterprises

■ **transport costs**

See trade costs and foreign direct investment

■ **transportation**

See air transportation; shipping

■ **Triffin dilemma**

Little more than ten years after the creation of the Bretton Woods system, which was founded on the belief that the gold standard had contributed to the financial instability and widespread unemployment in the 1920s and 1930s, the Yale University econo mist Robert Triffin predicted in two separate articles in 1959 60 that internal contradictions would soon bring to an end the new international architecture constructed at Bretton Woods. Around ten years later his prediction, commonly called the Triffin dilemma, proved to be correct.

The major problem with the gold standard system was the failure of gold supplies to keep pace with the requirements of an expanding international trading system for adequate reserves and liquidity and its asymmetric adjustment mechanism that led to a predominance of restrictive domestic policies and a tendency toward global recession. Since limited gold reserves were used to meet international payments deficits, a country would eventually have to restrict domestic demand for foreign goods by reducing economic activity. On the other hand, there was no necessity for countries with payments surpluses who were receiving gold to contribute to adjustment by increasing their imports. Thus there was always a bias toward restrictive policies.

Despite many proposals to base the new system on an international unit of account or an international currency whose supply could by managed by a su pranational bank to provide symmetric adjustment and support the needs of trade and global economic activity, the new system forged at the United Nations Monetary and Financial Conference held in 1944 at Bretton Woods, New Hampshire, was much less ambitious.

Under the Articles of Agreement of the Interna tional Monetary Fund (IMF), countries were given the choice of setting the par exchange value of their currencies for current account transactions in terms of gold or dollars. Since the United States held virtually all of the world's gold, it was the only country able to fix the par value of its currency in terms of gold. All other countries had no choice but to fix the par values of their currencies in terms of the dollar. This meant that these countries had to hold dollar reserves to ensure that market exchange rates did not diverge by more than 1.25 percent from par. The fixed exchange rate system was thus main tained by means of intervention in foreign ex change markets by all central banks, except the U.S. Federal Reserve, buying or selling dollars. The U.S. dollar was thus the sole intervention currency and effectively replaced gold in the Bretton Woods system.

Thus, the Bretton Woods system differed in one important respect from the gold standard. Instead of

the supply being determined by mining and dishoarding, the supply of dollars to serve as international liquidity was determined by a deficit on the external trade and financial payments balance of the United States. In the immediate postwar period this balance was in substantial surplus, and most experts believed that the main problem with the new system would be dollar scarcity and a lack of international liquidity.

U.S. Dollar Overhang Unilateral transfers by the United States for political and military purposes, however, eventually led to current account deficits and an increasing stock of dollars held outside the United States. Triffin pointed out that while the successful operation of this new international financial system depended on an expansion of the U.S. current account deficit to keep international liquidity growing in step with rapidly expanding world trade, this would eventually create foreign claims on the United States that exceeded the gold supply. Once foreign claims on the United States exceeded the dollar value of the gold supply at the $35-an-ounce parity, the United States could no longer guarantee convertibility of the dollar into gold at the official parity. Fixed exchange rates required that foreign claims on the United States be kept below its holdings of gold reserves, while maintaining the growth in the supply of global liquidity to meet the needs of trade required just the opposite.

The survival of the dollar-based Bretton Woods fixed exchange rate system then depended on the willingness of foreigners to refrain from buying gold with their dollar reserves. Foreign holders of the dollar thus faced a dilemma: if they converted excess dollar balances into gold they risked devaluation and losses on their dollar holdings measured in their domestic currency on their balance sheets, but if they delayed and the dollar was devalued relative to gold they risked even larger losses.

Many countries argued that the United States could resolve the problem of excess dollar balances if it reduced domestic growth to create balance of payments surpluses, but this remedy would simply have reduced the growth of international liquidity

and created the risk of global recession. It was just such a case that John Maynard Keynes had warned against when he pointed out the necessity of avoiding asymmetric balance of payments adjustments. During the 1960s the United States introduced a series of ad hoc measures to manage the convertibility of the dollar into gold and to increase foreign demand for dollars, but the internal contradictions elucidated by Triffin were not eliminated.

There were three ways to resolve the problem:

1. create an international currency issued by a global central bank to replace the dollar at the center of the international system,
2. produce global deflation through an attempt to reduce the supply of dollars to the level of the remaining U.S. gold supply, or
3. allow the U.S. to default on its foreign debts through the elimination of the convertibility of the dollar, devaluation of the dollar, and the introduction of flexible exchange rates.

Triffin's prediction was proven correct when the last remedy was chosen in the early 1970s and fixed exchange rates were abandoned. Abandoning the dollar peg to gold did not completely resolve the problems Triffin raised, however, nor did it eliminate asymmetric adjustment. The current international financial system seems to exhibit a variant of the Triffin dilemma due to divergent policy objectives between developed and developing countries. The former tend to pursue policies to ensure exchange rate and asset price stability and allocative efficiency of financial markets at the cost of domestic growth performance below potential, that is, conditions of excess supply. On the other hand, developing countries are more interested in increasing their income and wealth levels via rapid growth.

Post–Bretton Woods: Flexible Exchange Rates One advantage of the shift to a floating exchange rate system after the collapse of the Bretton Woods system was to make diverse policy objectives compatible. The free global flow of capital in search of the most remuneration should maximize global growth. Developing countries in the post Bretton Woods

world, however, have high potential growth rates relative to developed countries and need to run balance of payments deficits to fill resource gaps by importing necessary goods. To pay for these imports, the developing country must borrow from foreign suppliers, which is financing with foreign capital inflows.

These flows are supported by the assumption that developing countries with high growth potential also have high relative real rates of return on investment and thus are ideal investment targets for developed countries with excess savings and few attractive domestic outlets for investment. Large differentials in relative growth rates as well as large resource gaps require large capital inflows and associated balance of payments deficits. The larger the deficit, the higher the likelihood of an exchange rate adjustment or depreciation, creating higher volatility in returns and greater risks to developed country investors. Thus the more successful the system is in allowing developing countries to achieve their potential growth rates, the less willing developed country investors are to finance these deficits.

Thus the Triffin dilemma of the 21st century implies that the more successful developing countries are in achieving their potential growth rates by borrowing in international capital markets to fill their resource gaps, the larger will be their foreign imbalances and the less likely they will be to retain the capital inflows required to support the growth rate. Just as large U.S. deficits increased the risk of holding dollars, larger developing country deficits increase the risk of holding assets denominated in their currencies.

Under the Bretton Woods system, countries held dollar balances because central banks had to support their Bretton Woods parity, and the dollar retained international value in excess of its gold backing. Since the Bretton Woods system ended, exchange rates in developing countries are held up by the force of capital flows themselves, which prior to crises exceed the funds needed to meet balance of payments shortfalls and produce a continuous upward pressure on the currency. Also, because of their higher po-

tential growth rates, the monetary policies in most developing countries produce interest rates that are higher than those abroad, providing an additional attraction for foreign investors. This produces a condition for foreign investors that is much like the catch-22 faced by central banks under Bretton Woods. As the balance of payments continues to deteriorate, the required risk-adjusted rate of return on investment increases but withdrawing capital will inevitably produce exchange rate instability that erodes or eliminates the nominal excess returns expected to be earned in developing countries. Eventually the balance tips toward increased risk, and an exchange rate crisis requires an adjustment policy that generates a decline in income growth and employment.

Thus the same asymmetric adjustment that made deficit countries adjust their policies to those of surplus countries also forces developing countries to make their policies compatible with the policy objectives of the developed countries. Post Bretton Woods, it is the reversal of international capital flows that produces the adjustment. A country that grows too rapidly and produces a large balance of payments deficit that threatens exchange rate stability, and thus its excess rate of return, finds that foreign capital flows out and a financial crisis forces adjustment.

New Triffin Dilemma After the 1997 Asian crisis, many developing countries shifted their policies toward current account surpluses and the accumulation of dollar reserves to protect themselves against capital outflows and the need to borrow from the IMF. As a result, the United States has accumulated large external account imbalances, so dollar holders face the same Triffin dilemma that occurred at the end of the Bretton Woods system. The Triffin dilemma thus seems to be of more general application than its author might have originally envisaged.

See also balance of payments; Bretton Woods system; convertibility; currency crisis; dollar standard; exchange rate regimes; financial crisis; global imbalances; gold

standard, international; international liquidity; International Monetary Fund (IMF); International Monetary Fund conditionality; international reserves; money supply; twin deficits

FURTHER READING

International Monetary Arrangements: The Problem of Choice Report on the Deliberations of an International Study Group of 32 Economists. 1964. Princeton, NJ: Princeton University International Finance Section. Proposals from an eminent group of international monetary experts on possible alternative international monetary arrangements.

Lary, Hal B. 1963. *Problems of the United States as World Trader and Banker.* New York: Columbia University Press for the National Bureau of Economic Research. Classic study outlining the conflicting role of the United States as a trading nation and as the provider of the global means of exchange.

Triffin, Robert. 1960. *Gold and the Dollar Crisis.* New Haven, CT: Yale University Press. Reprints of Triffin's original articles enunciating the dilemma.

———. 1966. *The World Money Maze: National Currencies in International Payments.* New Haven, CT: Yale University Press. Further extensions on the problems caused by an international monetary system based on national currencies.

JAN KREGEL

■ twin deficits

The *twin deficits hypothesis* refers to the belief that changes in fiscal policy that increase the budget deficit will also increase the current account deficit. The term gained popularity in the 1980s, when tax cuts that caused the U.S. budget deficit to increase by 3 percent of gross domestic product (GDP) between 1980 and 1985 were followed by trade deficits exceeding 3 percent of GDP in 1985 and 1986. The concept received renewed attention in the first decade of the 21st century, when a deterioration of the U.S. long-term or structural budget balance of 6 percent of GDP between 2000 and 2004 was accompanied by current account deficits of 5 percent of GDP in 2004 and 6.5 percent of GDP in 2005.

Accounting and Theoretical Relations In terms of accounting, the current account deficit and the budget deficit are related through the identity:

$$\text{Saving} + \text{Current Account Deficit}$$
$$= \text{Investment} + \text{Budget Deficit} \qquad (1)$$

The identity implies that an increase in the budget deficit that does not cause a change in private saving or investment will increase the current account deficit. Equation (1) can be rewritten as:

$$\text{Current Account Deficit}$$
$$= \text{Investment} - \text{National Saving} \qquad (2)$$

where national saving equals private saving minus the budget deficit (or private saving plus government saving). Written in this way, the identity highlights the fact that a country with insufficient domestic saving to finance its domestic investment must make up the shortfall by running a current account deficit (i.e., drawing on foreign saving).

To understand the relationship between budget deficits and current account deficits it is necessary to go beyond accounting identities to theoretical models. In the Mundell-Fleming model, a fiscal expansion increases aggregate demand and the domestic interest rate. This in turn generates a capital inflow that causes the currency to appreciate. The stronger currency then reduces net exports, returning aggregate demand and output to their original levels. Similar transmission mechanisms are operative in larger-scale general equilibrium macro models that build on equation (2) and the fact that income and exchange rate changes affect exports and imports (e.g., see Cline 2005).

An alternative theoretical perspective in which budget deficits do not affect current account deficits is the Ricardian equivalence model (Barro 1974). Under this theory, a reduction in taxes that raises the budget deficit will be followed by an increase in taxes in the future. Rational taxpayers will anticipate these future tax liabilities that will fall either on them or on their descendants and will increase savings by the amount of the tax cuts. Thus in equation (1), savings will increase by the same amount as the budget deficit. An increase in the budget deficit will thus have no effect on in-

terest rates, exchange rates, investment, or net exports.

Choosing empirically between the Ricardian equivalence model and models where deficits matter has proven difficult. The simplest test may be to examine whether budget deficits raise interest rates. Since budget deficits vary countercyclically and interest rates vary cyclically, however, these tests are biased against finding a relationship between the variables. After many empirical studies of the Ricardian model, no clear consensus has emerged.

In spite of the difficulty with empirical testing, many have argued that the theoretical requirements for Ricardian equivalence to hold are too stringent (see, e.g., Haliassos and Tobin 1990). The model requires parents, children, grandchildren, great-grandchildren, and so on to be linked in an unbroken chain through the giving of intergenerational gifts. It requires capital markets to be perfect and agents to be able to borrow and lend on the same terms as the government. It requires taxes and transfers to be lump sum. Finally, it requires that tax cuts be followed by tax increases rather than by new debt issues to service the existing debt. Because these theoretical requirements seem too restrictive, many have argued that an increase in the budget deficit will not produce a dollar-for-dollar increase in private saving. In this case, an increase in the budget deficit will partly crowd out either net exports (the twin deficits hypothesis) or private investment.

The Mainstream Perspective on the Twin Deficits Proponents of the twin deficits hypothesis do not argue that budget deficits and current account deficits move in lockstep, but only that fiscal policy changes that increase the budget deficit can also increase the trade deficit. Many factors can attenuate the relationship between the two variables. For instance, a stock market boom can reduce savings, increase investment, and attract funds from abroad. These factors would all lead to a larger current account deficit. Greater tax revenues from the stock market boom, however, would move the budget toward surplus. Similarly a recession would reduce imports and improve the trade balance but reduce tax revenues and worsen the budget deficit.

The fact that the two deficits do not always move together is borne out by the data. The two deficits moved in the same direction in only ten years between 1980 and 2004 (Cline 2005). Formal econometric and calibration studies indicate that a $1 reduction in the budget deficit will reduce the current account deficit by between 20 and 40 cents (see Erceg, Guerrieri, and Gust 2005; Cline 2005; Chinn and Prasad 2003).

A major concern with the twin deficits is whether the borrowing associated with them will prove sustainable. The same framework can be used to think about the growth of government debt and the growth of external debt (Gramlich 2004). In both cases, the long-run value of the stock of debt relative to GDP n^* is given by:

$$n^* = d/(g - r) \qquad (3)$$

where d is the nominal primary budget deficit or current account deficit relative to GDP, g is the nominal growth rate of the economy, and r is the net nominal interest rate on the debt. Many authors have calculated that, given current trends, the stock of U.S. external debt could reach levels exceeding 100 percent of U.S. GDP.

An important question is whether foreign investors will at some point become unwilling to accumulate additional U.S. debt. This could lead to a hard landing for the dollar and the U.S. economy. The reluctance of foreigners to acquire additional U.S. assets could cause the dollar to fall, interest rates to rise, and stock prices to tumble. The adjustment could prove disorderly if the fall of the dollar led to a rush for the exits among investors holding dollar assets.

Arguments That the Twin Deficits Do Not Matter While many are concerned that the twin deficits will cause a hard landing for the U.S. economy, others are more sanguine. Arguments that the twin deficits do not matter point to valuation effects, the size of global capital markets, the impact of a global savings glut, and China's willingness to accumulate dollar assets.

Valuation effects refers to factors that offset the effect of large current account deficits on the value of the U.S. external debt. One such factor can be seen from equation (3). The interest rate r has been per-

sistently negative (i.e., favorable to the United States), driving down the value of n*. The interest rate is negative because, up until now, investment income earned by U.S. residents from abroad has exceeded investment income paid by U.S. residents to the rest of the world. Since net international investment income has remained positive for the United States, some have in fact argued that the United States is a creditor and not a debtor in an economic sense (Cline 2005).

In a similar vein the economists Hausmann and Sturzenegger (2006) have argued that positive net income for the United States in every year from 1980 to 2005 implies that the economic value of U.S.-owned assets abroad exceeds the economic value of foreign-owned assets in the United States. They attribute the difference between the U.S. economic status as a creditor and the official data showing that the United States is a debtor to the presence of "dark matter." *Dark matter* refers to the value of know-how and brand recognition associated with U.S. foreign direct investment (FDI) abroad, the seigniorage gains the United States receives because people throughout the world are willing to hold U.S. dollars, and the "insurance premium" that the United States receives because investors are willing to hold U.S. Treasury securities at a lower rate of return than the United States can earn on foreign currency assets.

Others take issue with the argument that, because of dark matter, the United States is actually a net creditor. The data on net international investment income may be inaccurate because foreign firms underreport profits from U.S. operations to minimize tax liabilities. Further, U.S. assets from abroad are largely in the form of equities (including FDI), while U.S. external liabilities are largely in the form of fixed-income assets (including U.S. Treasury securities). Thus the average return received by U.S. investors exceeds the average return received by foreign investors because the United States is holding riskier and more profitable assets (Frankel 2006).

Some have argued that because global capital markets are large, the United States can continue borrowing massive amounts in the future without the

U.S. economy experiencing a hard landing (Cooper 2005). According to one calculation, the non-U.S. financial wealth equals about 700 percent of U.S. GDP (Cline 2005). Assuming that "home bias" in the rest of the world continues to decline, some have argued that foreign investors in the future may be willing to hold even the very large quantities of U.S. external debt implied by large U.S. current account deficits. Others suggest that country risk is better measured by a country's debt relative to its own exports or GDP rather than relative to the size of foreign portfolios (Frankel 2006). If measured by these criteria, U.S. external debt is on a trajectory in the first decade of the 21st century to reach precarious levels (see Obtsfeld and Rogoff 2005). Many argue that this growth in U.S. debt/GDP and U.S. debt/exports will multiply the risk of a hard landing.

The chairman of the Federal Reserve, Ben Bernanke (2005), has argued that the budget deficits of the George W. Bush administration were not an important cause of large trade deficits but instead that the external imbalances were driven by a global savings glut. Indeed many developing and developed countries have excesses of saving over investment that have flowed to the United States. These surplus funds have in turn reduced U.S. interest rates and pushed up U.S. housing prices. The resulting increase in household wealth has reduced U.S. private saving and led to large trade deficits. According to calculations by Cline (2005), a global savings glut can explain at most 35 percent of the U.S. current account deficit. The rest he attributes to U.S. fiscal policy and other factors.

Another argument against the significance of the twin deficits is that China and other developing countries running surpluses with the United States are willing to accumulate massive quantities of dollar assets. It has been suggested that this arrangement constitutes a new Bretton Woods system (see Dooley, Folkerts-Landau, and Garber 2004). China and East Asia are playing a role analogous to the role played by Europe in the 1960s. China has accumulated $1 trillion of reserves, predominantly in dollars, in order to maintain a competitive exchange rate and thus sustain the export-oriented thrust of its econ-

omy. This strategy will help China to provide jobs for hundreds of millions of underemployed rural workers and also facilitate the development of financial infrastructure and modern systems of corporate governance. If this thesis is accepted, the current system could well be sustained for a long time.

Others claim that this arrangement is not sustainable. For instance, China may either develop a workable system of finance and corporate governance or experience a financial crisis (Frankel 2006). Either way, the excess liquidity flowing from China to the United States will stop. In addition, foreign reserve accumulation can increase the money supply and exacerbate inflation. Sterilization policies to offset this effect may force commercial banks to hold more and more central bank bills, erode bank profitability, and interfere with the allocation of credit through the banking system. Further, continued accumulation of U.S. Treasury securities (external reserves) results in an increasingly inefficient allocation of resources since both private and social rates of return would be much higher for investments in the domestic economy (see Summers 2006). Thus many believe that foreign reserve accumulation by Chinese and other central banks will not prove sustainable.

There is thus a lively debate about the empirical relevance of the twin deficits hypothesis. More research is needed to clarify the relationships among budget deficits, current account deficits, and other variables.

See also balance of payments; Bretton Woods system; Federal Reserve Board; foreign exchange intervention; global imbalances; home country bias; international financial architecture; international reserves; mercantilism; Mundell-Fleming model; seigniorage; sterilization

FURTHER READING

Barro, Robert. 1974. "Are Government Bonds Net Wealth?" *Journal of Political Economy* 82: 1095 1117. Indicates that budget deficits do not affect current account deficits.

Bernanke, Ben. 2005. "The Global Saving Glut and the U.S. Current Account Deficit." Lecture presented to the Virginia Association of Economics, Richmond, VA, April 14, 2005. Downloadable at http://www .federalreserve.gov/boarddocs/speeches/2005/default .htm. Argues that America's current account deficits in the 21st century are primarily due to excess saving abroad rather than U.S. budget deficits.

Chinn, Menzie, and Eswar Prasad. 2003. "Medium Term Determinants of Current Accounts in Industrial and Developing Countries: An Empirical Exploration." *Journal of International Economics* 59 (1): 47 76. Investigates the link between budget deficits and current account deficits.

Cline, William. 2005. *The United States as a Debtor Nation.* Washington, DC: Institute for International Economics. A comprehensive and extraordinarily useful reference on the subject.

Cooper, Richard. 2005. "Living with Global Imbalances: A Contrarian View." Institute for International Economics Policy Brief 05 3. Washington, DC: Institute for International Economics. Argues that, because global capital markets are large, the United States can continue borrowing massive amounts in the future without the U.S. economy experiencing a hard landing

Dooley, Michael P., David Folkerts Landau, and Peter Garber. 2004. "The Revised Bretton Woods System." *International Journal of Finance and Economics* 9 (4): 307 13. Argues that China and East Asia will continue to accumulate dollar reserves for a long time.

Erceg, Christopher, Luca Guerrieri, and Christopher Gust. 2005. "Expansionary Fiscal Policy and the Trade Deficit." Federal Reserve International Finance Discussion Paper 2005 825. Washington, DC: Federal Reserve Board. Provides quantitative evidence of the link between budget deficits and current account deficits.

Frankel, Jeffrey. 2006. "Could the Twin Deficits Jeopardize U.S. Hegemony." *Journal of Policy Modeling* 28 (6): 653 63. An excellent defense of the position that the twin deficits matter.

Gramlich, Edward. 2004. "Budget and Trade Deficits: Linked, Both Worrisome in the Long Run, but Not Twins." Washington, DC: Federal Reserve Board. http://www.federalreserve.gov/boarddocs/speeches/ 2004/20040225/default.htm. A clear discussion of whether the borrowing associated with the twin deficits is sustainable.

Haliassos, Michael, and James Tobin. 1990. "The Macroeconomics of Government Finance." In *Handbook of*

Monetary Economics, edited by Benjamin Friedman and Frank Hahn. New York: Elsevier Science, 889 959. An excellent and comprehensive discussion of the macroeconomics of government borrowing.

Hausmann, Ricardo, and Federico Sturzenegger. 2006. "Global Imbalances or Bad Accounting?: The Missing Dark Matter in the Wealth of Nations." Kennedy School of Government Working Paper RWP 06 003. Cambridge, MA: Harvard University. Argues controversially that the United States is actually a creditor nation rather than a debtor nation.

Obtsfeld, Maurice, and Kenneth Rogoff. 2005. "The Unsustainable U.S. Current Account Position Revisited." In *G7 Current Account Imbalances: Sustainability and Adjustment*, edited by Richard Clarida. Chicago: University of Chicago Press, 339 66. Argues that the U.S. current account deficit is unsustainable and poses risks to the world economy.

Summers, Lawrence. 2006. "Reflections on Global Account Imbalances and Emerging Markets Reserve Accumulation." Speech presented at the Reserve Bank of India (March 24). Downloadable at http://www.president.harvard.edu/speeches/summers/2006/0324 rbi.html. Argues that central banks should hold external reserves in assets other than U.S. Treasury securities.

WILLEM THORBECKE

U

■ unions and foreign direct investment

Unionization has been falling over time, especially in the United States. In 2005, the unionization rate in Great Britain was 29 percent; Japan, 18.7 percent; Korea, 11.9 percent; and the United States, 12.5 percent. According to data for other years, the unionization rate in Australia was 24.7 percent (2000 data); Canada, 29.7 percent (2006 data), Germany, 26.6 percent (2002 data); Singapore, 21.5 percent (2004 data); and the Philippines, 26.0 percent (2002 data). Meanwhile foreign direct investment (FDI) has been growing rapidly. The United States, the largest recipient of FDI, has one of the lowest unionization rates. Is there a connection?

Theoretical Frameworks FDI can benefit or hurt labor, depending on factors such as its motivations, the structure of firms, the organization of labor-management negotiations, the preference of labor unions, and the location of workers. Firms can be horizontally or vertically related; labor-management bargaining can be firm specific, industrywide, or even nationally centralized; and unions may prefer wages more than employment, or vice versa.

Often the motivation for undertaking FDI is market access. With the classical *tariff-jumping FDI*, the motivation is to jump over existing import barriers such as tariffs, quotas, anti-dumping duties, and safeguard measures. By producing directly in the destination country, the multinational firm avoids import barriers that a national firm must face when exporting.

Related is the so-called *quid pro quo FDI*. The quid pro quo literature suggests that FDI may be induced by the threat of protection (possible protection in the future), and furthermore, that FDI may be used as an instrument to defuse a future protectionist threat. By investing directly in the host countries, foreign firms are creating jobs in the host country. As a consequence, domestic lobbying pressure against foreign imports decreases.

If FDI is undertaken for the purposes just mentioned, it benefits labor in both countries compared with the case of no FDI at all, because it raises output and employment, which in turn puts upward pressure on wages.

In unionized industries, multinationals may undertake FDI to reduce labor costs, to improve their relative bargaining position with the union, or to avoid employing union workers. Consequently, unions historically tend to view multinationals and FDI negatively, regardless of their sources or destinations.

When the motive is to reduce labor costs, FDI may benefit or hurt labor, depending on whether the multinational enterprise (MNE) is horizontally or vertically structured. A horizontal MNE produces identical goods in several countries, while a vertical MNE produces inputs in some countries and final goods in other countries. The theoretical literature suggests that horizontal MNE can hurt labor, because by moving production facilities across borders, firms shift the demand for labor, which drives down the negotiated union wage (see for instance Zhao 2001 and the references cited). Further, usually labor unions are assumed to maximize an objective function consisting of union employment and a union

wage premium above the nonunion wage. If the labor unions are more interested in higher employment than in the wage premium, then horizontal FDI reduces union employment also. This result remains true regardless of whether labor-management bargaining is firm specific or industrywide. And if labor-management bargaining is firm specific and unionization is industrywide, the previously mentioned effects of FDI are substantially reduced, because union members can find work in other firms if bargaining breaks down. In this case, unions may welcome horizontal FDI also. In addition, under subsidiary bargaining in which the union negotiates with branches (not headquarters) of the MNE, union strikes affect only portions of the MNE earnings, while under headquarter bargaining in which the union negotiates with the MNE headquarters, the MNE's total earnings are affected. Thus the labor union prefers to bargain with the headquarters rather than with the subsidiary, even though the MNE may use transfer pricing to strategically influence the negotiated wages and employment by shifting profits across branches to gain a better position in negotiations.

Leahy and Montagna (2000) investigate the effects of different degrees of centralization in wage setting on the incentive of a MNE to locate in a host country, on the host country's welfare, and on the MNE's preference for centralized or decentralized bargaining. They find that if international product markets are interrelated, then centralized bargaining would help the MNE and hurt the host country. This arises because the more efficient MNE would try to raise the wage in the host country above the level host-country firms can afford (i.e., raising rivals' costs), thereby capturing a larger market share. Skaksen and Sorensen (2001) show that home workers may lose or gain depending on the substitutability of the multinational activities. If there is a high degree of substitutability (complementariness) between the activities in the home country and the host country, it is likely that the workers lose (gain) on FDI. Naylor and Santoni (2003) show that FDI is less likely, other things being equal, the greater the union bargaining power and the more substitutable the firms' products in the potential host country.

On the other hand, a vertical MNE can be good or bad for labor in the source country, depending on host-country wages and how much production is shifted overseas. Zhao and Okamura (2008) show that undertaking FDI in a lower-wage country has two effects: an output-expansion effect and a demand-shifting effect. The output-expansion effect stems from using cheaper labor in the host country, which reduces the MNE's overall average cost and enables higher output and employment. The demand-shifting effect is that FDI reduces labor demand in the source country but increases that in the host country. If the first effect dominates the second, then FDI benefits labor in the source country as well as in the host country. This case arises more often if the wage in the host country is not too much below that in the source country. Otherwise the second effect dominates the first and labor in the source country loses from FDI. One important conclusion is that if the wage in the host country is relatively high (nevertheless lower than in the source country), then FDI benefits the labor union and the MNE in the source country, as well as the host country itself.

Empirical Studies There are very few empirical studies concerning labor unions and FDI, and they use U.S. data and mostly confirm the theoretical predictions. Using U.S. Department of Commerce data published in 1992, Cooke (1997) examines the influence of several key industrial relations variables on U.S. outward FDI across nine industries and 19 members of the Organisation for Economic Co-operation and Development. He finds that FDI is negatively related to high levels of unionization, centralized bargaining structures, and industrial relations environments unfavorable to firms such as governmental restrictions on layoffs, but FDI is positively affected by high levels of worker education and policies requiring works councils. Using 1981–83 data on U.S. inward FDI, however, Coughlin, Terza, and Arromdee (1991) are surprised to find that higher unionization rates were associated with increased FDI.

Slaughter (2007) examines whether globalization has played a role in the falling private-sector unionization rate in the United States, with unions feeling pressured to reduce employment and/or compensation demands. He assembles a panel of U.S. manufacturing industries that matches union-coverage rates with measures of global engagement such as exports, imports, tariffs, transportation costs, and FDI. He finds a statistically and economically significant correlation between falling union coverage and greater numbers of inward FDI transactions. Because U.S. affiliates of foreign multinationals have higher unionization rates than U.S.-based firms do, this correlation may reflect pressure of international capital mobility on U.S.-based companies, consistent with research on how rising capital mobility raises labor-demand elasticities and alters bargaining power.

In sum, while the theories of unions and FDI are many, the empirical tests of these theories are few. More work is needed to establish the effects of FDI on unions, of unions on FDI, and of unions on how FDI affects host and source countries.

See also factor endowments and foreign direct investment; footloose production; foreign direct investment and labor markets; technology licensing; trade and wages

FURTHER READING

Cooke, William N. 1997. "The Influence of Industrial Relations Factors on U.S. Foreign Direct Investment Abroad." *Industrial and Labor Relations Review* 51 (1): 3 17. Finds FDI to be negatively related to unionization.

Coughlin, Cletis C., Joseph V. Terza, and Vachira Arromdee. 1991. "State Characteristics and the Location of Foreign Direct Investment within the United States." *Review of Economics and Statistics* 73 (4): 675 83. Finds unionization attracts FDI using U.S. data.

Leahy, Dermot, and Catia Montagna. 2000. " Unionization and Foreign Direct Investment: Challenging Conventional Wisdom?" *Economic Journal* 110 (462): C80 92. Shows that centralized bargaining helps the MNE but hurts the host country.

Naylor, Robin, and Michele Santoni. 2003. "Foreign Direct Investment and Wage Bargaining." *Journal of Interna tional Trade and Economic Development* 12 (1): 1 18. Shows that FDI is less likely with greater union bar gaining power or more substitutable products.

Skaksen, Mette Yde, and Jan Rose Sorensen. 2001. "Should Trade Unions Appreciate Foreign Direct Investment." *Journal of International Economics* 55 (2): 379 90. Shows that workers may lose or gain depending on the substitutability of multinational activities.

Slaughter, Matthew J. 2007. "Globalization and Declining Unionization in the United States." *Industrial Relations* 46 (2): 329 46. Finds significant correlation between falling union coverage and rising inward FDI.

Zhao, Laixun. 2001. "Unionization, Vertical Markets, and the Outsourcing of Multinationals." *Journal of Interna tional Economics* 55: 187 202. Analyzes horizontal and vertical MNE under unionization. References include many other contributions on union and FDI.

Zhao, Laixun, and Makoto Okamura. 2008. "Competing to Invest in the Foreign Market." Working Paper No. 217. Research Institute for Economics & Business, Kobe University. Presents cases of FDI benefiting the labor union, the MNE, as well as the host country.

LAIXUN ZHAO

■ United Nations Conference on Trade and Development

The United Nations Conference on Trade and Development (UNCTAD) was set up in 1964 as a forum for the negotiation of international trade policy issues that fell outside the scope of the General Agreement on Tariffs and Trade (GATT). In the 1990s, with the establishment of the World Trade Organization (WTO), it lost that function and has since become the UN body responsible for research, advocacy, and technical assistance on international trade issues.

The seeds of UNCTAD lie in the failure to establish an International Trade Organization (ITO) after World War II. Although negotiations for an ITO were concluded successfully in 1947 48, the United States subsequently failed to ratify the agreement and the ITO was stillborn. All that remained

was an interim GATT, which covered the process for negotiating reduced tariffs on industrial goods traded across borders. This was of limited interest to developing countries that were then scarcely industrialized. Their major concern was the regulation of international trade in primary commodities, which had been provided for, to a limited extent, in the failed ITO agreement. During the 1950s, they pressed in the forum of the United Nations for new international machinery to address this and other issues of special interest to them.

Geopolitical circumstances proved favorable to their demands. As part of its Cold War tactics, the Soviet Union supported the developing countries' demand for a new international trade conference. The John F. Kennedy administration decided to abandon the U.S. government's public opposition to that proposal to prevent the issue from recruiting developing countries into the Soviet camp. As a result, the first UN Conference on Trade and Development was convened in Geneva from March to June 1964. The secretary-general was Raúl Prebisch, an Argentine economist who had previously been executive secretary of the UN Economic Commission for Latin America. The developing countries organized themselves into a caucus known as the Group of 77   even though its membership eventually exceeded 100.

Little was achieved initially, and the continued existence of the conference was at first in doubt, but eventually the United States agreed that it should reconvene. Thus UNCTAD is unlike most other international organizations. It is an international conference convened every four years in a different location and is serviced by a section of the UN Secretariat, based mainly in Geneva and numbering about 440 officials. Its governing organ, the Trade and Development Committee, meets regularly in Geneva between sessions of the conference to direct the Secretariat's programs.

Campaign for International Commodity Agreements The first several conferences were the site of negotiations on various trade issues that were not being negotiated in the GATT, but on which the UNCTAD Secretariat had produced a range of in-terventionist proposals. Most important, in 1974 UNCTAD emerged to international prominence as one of the forums for globally negotiating a New International Economic Order (NIEO). These negotiations centered on the Secretariat's proposal to extend the number of International Commodity Agreements (ICAs). The aim was to dampen the volatility of primary commodity prices, using a buffer stock mechanism, to be financed by a Common Fund that would also fund commodity diversification schemes.

This package of policies ultimately proved insufficiently attractive for the commodity-consuming countries, which did not really want a fund that they did not control and were more concerned about the security of oil supplies than security of supply of other primary commodities. The commodity-producing countries were also ultimately   for quite different reasons   ambivalent about it. Since many developing countries were highly dependent on just a few commodities, each was worried that a new Common Fund would weaken the control that they already exercised within existing ICAs over the commodities of special interest to them. In the end, no new ICAs emerged and the Common Fund that came into being was very much smaller than the $6 billion Fund first envisaged by Gamani Corea, UNCTAD's secretary-general from 1974 to 1980.

Post-NIEO Development UNCTAD has taken up a variety of other trade-related issues of special concern to developing countries. The transfer of technology was a concern that in the 1980s led to attempts to negotiate an international code of conduct on the subject. Projects of economic cooperation among developing countries were taken up as a theme in UNCTAD, as well as special issues affecting the least-developed countries. The results of this proliferation of activity following the NIEO failure were varied. It contributed to a loss of coherence and focus in the organization as a whole, as small units engaged on marginal and unrelated work multiplied. This fragmentation was aggravated by the UNCTAD Secretariat's growing practice of taking on technical assistance work as a subcontractor to other international public agencies. On the other side

of the account, in some of these new niche activities, the UNCTAD Secretariat was able to develop special skills and deliver technical assistance that was valuable. By the 1990s, its sternest critics recognized its achievements in several important areas the provision of information on statistics of international trade and of global data on private foreign investment and specialized technical assistance in debt management systems.

UNCTAD has also maintained some high-quality publications that regularly analyze the world economic situation and its impact on trade, finance, and development. The annual *Trade and Development Report* series is preeminent in that respect. The UNCTAD Secretariat publishes several flagship reports, the others being the *World Investment Report* and *Least Developed Countries Report.*

Impact of the World Trade Organization The establishment in 1995 of the World Trade Organization (WTO) radically changed UNCTAD's position among the international trade institutions. In the presence of the WTO, it no longer made sense to have two international bodies for the negotiation of trade issues. The disempowerment of UNCTAD as a North-South negotiating forum started at UNCTAD VIII in Cartagena (1992) and was completed at UNCTAD IX at Midrand, South Africa (1996). After 1996, UNCTAD ceased to be a negotiating forum for world trade issues, and in line with this, it reduced the scale of its intergovernmental machinery and consolidated its divisional structure. It replaced 25 separate work programs and subprograms with 1 program consisting of 5 subprograms; it halved the number of intergovernmental bodies; it cut the number of meetings to one-third of what it had been in 1992; and it reduced the number of divisions in the secretariat from 9 to 5.

From 1996 on, its role has been ancillary to the WTO, providing a place to discuss trade and development issues, advocating for the views of developing countries, and offering advice and technical assistance to developing countries preparing for trade negotiations in the WTO. This reduction in function and size implied greater interagency cooperation

between UNCTAD and the WTO, building on previous GATT/UNCTAD cooperation, such as their joint International Trade Center to help developing countries to promote their exports.

The duration of each full session of UNCTAD has been shortened. The 11th session, held in July 2004 in Rio de Janeiro, lasted only a week and was used as a showcase for the organization and its remaining functions. As UNCTAD secretary-general from 1995 to 2005, Rubens Ricupero, himself a former trade negotiator for Brazil, was successful in adapting UNCTAD to this more modest role, and thereby gaining the goodwill of G8 governments. As of 2008, the current secretary-general is Panitchpakdi Supachai of Thailand, who was director-general of the WTO from 2002 to 2005.

No longer a forum for major global trade negotiations, UNCTAD is confined to carrying out international trade research, advocacy, and technical assistance functions and acting as a junior partner of the WTO.

See also trade and economic development, international; World Trade Organization

FURTHER READING

Corea, Gamani. 1992. *Taming Commodity Markets: The Integrated Program and the Common Fund in UNCTAD.* Manchester, UK: Manchester University Press. A history of the NIEO negotiation by a former UNCTAD secretary general.

Griffin, Keith. 2003. "Economic Globalization and Institutions of Global Governance." *Development and Change* 34 (5): 789 808. Places UNCTAD within the spectrum of global governance institutions.

Henderson, David. 1998. "International Agencies and Cross Border Liberalization." In *The WTO as an International Organization*, edited by Anne Krueger. Chicago: University of Chicago Press, 97 130. A critical and comparative evaluation of international bodies concerned with trade.

Toye, Richard. 2003. "Developing Multilateralism: The Havana Charter and the Fight for the International Trade Organization, 1947 48." *International History Review* 25 (June): 282 305. An account of the negotiations around the failed ITO and GATT.

Toye, John, and Richard Toye. 2004. *The UN and Global Political Economy*. Bloomington: Indiana University Press. A recent history of the role of economists within UNCTAD.

United Nations. 1985. *The History of UNCTAD: 1964 84*. Geneva: UNCTAD. An in house 20 year history of UNCTAD.

———. 2004. *Beyond Conventional Wisdom in Development Policy: An Intellectual History of UNCTAD, 1964 2004*. Geneva: UNCTAD. An in house 40 year intellectual history of UNCTAD.

JOHN TOYE

■ Uruguay Round

The Uruguay Round of multilateral trade negotiations (1986 94) was launched during a period when the world economy was experiencing slow growth. Stagflation a mix of rising prices, weak output growth, and rising unemployment was complemented by a debt crisis that affected a number of major developing countries. The need for developing countries to generate foreign exchange to service debt, the rapid growth of export capacity for labor-intensive manufactures such as textiles and clothing in East Asian economies (the "newly industrializing countries" or "East Asian tigers"), and rising trade imbalances gave rise to strong pressures to restrict imports in major countries of the Organisation for Economic Co-operation and Development (OECD).

Although large-scale protectionism was avoided during the late 1970s and early 1980s, extensive recourse was made to quantitative import restrictions, "voluntary" export restraint agreements (VERs), and anti-dumping actions. Trade restrictions formed part of an inappropriate policy response to structural adjustment pressures. Starting in the early 1980s, a number of contracting parties to the General Agreement on Tariffs and Trade (GATT) suggested that renewed efforts were needed to strengthen the trading system to address both rising protectionist pressures and to subject key areas of trade policy to multilateral disciplines. Of central importance

was agriculture largely exempted from GATT disciplines and the Multifiber Arrangement (MFA), a complex system of bilateral trade restrictions that had been put in place to manage trade in textiles and clothing. In addition the United States argued for expanding the coverage of multilateral rules to new areas such as trade in services, intellectual property rights, and investment.

The Uruguay Round was the eighth round of trade negotiations under GATT auspices. It was very ambitious, spanning many more subjects than any of its predecessors or its successor, the Doha Round. Among the major achievements of the Uruguay Round were the creation of the World Trade Organization, new multilateral disciplines for trade in services and intellectual property, and the reintegration of agriculture into the trade regime. It also led to further liberalization of international trade in merchandise. In addition to an average 40 percent cut in tariff bindings, there was agreement to abolish quantitative restrictions on textiles and clothing trade over a 10-year period and, for the first time, to impose specific limits on agricultural export subsidies and trade-distorting production support. Few observers at the time had expected such an ambitious outcome to be feasible. At numerous points during the eight-year marathon the talks came close to failure. Their successful conclusion marks a high point for multilateral cooperation in the 20th century.

The process that led to the establishment of the negotiating agenda for the Uruguay Round took five years, starting with the preparation for a 1982 ministerial meeting on trade, during which the United States sought but failed to obtain agreement to launch a new round. This failure was not surprising as the previous round, the Tokyo Round, had only recently been concluded. The work program that was agreed at the 1982 ministerial laid the foundation for the Uruguay Round, however. Working groups were established to review, among other topics, the use of quantitative restrictions, agricultural trade policies, and trade in services. In March 1985 a study group established by Arthur Dunkel, Director-General of the GATT, and chaired by Fritz Leutwiler, the president of the Swiss central bank, who had been

actively involved in efforts to address the ongoing debt crisis, issued the Leutwiler report. This provided support for the launch of new negotiations to tackle trade restrictions in agriculture and labor-intensive manufactures.

The 18 months after the publication of the Leutwiler report saw active efforts by proponents and opponents of a new round. The former spanned the OECD countries and many developing countries, especially in East Asia, that confronted trade restrictions in export markets. The latter comprised a number of developing countries, including India. By the time of the 1986 ministerial meeting in Punta del Este, Uruguay, the majority of the GATT contracting parties supported launch of a new round of negotiations. The main opponents a group of ten developing countries (the G10, which included Argentina, Brazil, Cuba, Egypt, India, Nicaragua, Nigeria, Peru, Tanzania, and Yugoslavia) sought to make new negotiations conditional on so-called standstill and rollback of existing trade restrictions imposed by OECD nations. They also rejected launching talks on services, trade-related aspects of intellectual property rights (IPRs), and trade-related investment measures (TRIMs).

The majority prevailed, adopting a compromise draft text proposed by Colombia and Switzerland. The major concession to the G10 was agreement that the services negotiations would proceed on a parallel track to talks on goods and that disagreements on IPRs and TRIMs would not be allowed to affect progress in other areas. Notwithstanding this procedural agreement, it was agreed that the negotiations were to be a "single undertaking."

The Negotiations: 1986–93 The talks spanned 14 areas related to trade in goods, plus services. The 14 subjects were tariffs, nontariff measures, tropical products, natural resource based products, textiles and clothing, agriculture, GATT articles (rules), safeguards, the codes of conduct negotiated during the Tokyo Round, subsidies and countervailing measures, dispute settlement, trade-related aspects of IPRs, TRIMs, and "the functioning of the trading system." It was agreed that the talks were to be concluded by the end of 1990

and that there would be a midterm review meeting in the course of 1988.

The midterm review was held in Montreal in December 1988. Differences on agriculture dominated the meeting, but there was disagreement on other subjects as well, including liberalization of trade in textiles and clothing, how to deal with safeguard actions and VERs, and the scope of ambition regarding protection of intellectual property. Instead of an "early harvest" of agreements in a variety of areas, the meeting failed and the midterm review was suspended. The primary reason was the huge disparity in objectives on agriculture between the United States and the EU, with the former insisting that complete liberalization should be the long-term goal. Given that no compromise seemed possible, the main protagonists suggested that ministers sign off on interim agreements that had been reached in other areas and instruct negotiators to return to Geneva and continue efforts to negotiate on agriculture. This "solution" to the deadlock was rejected by a number of Latin American countries. For the first time in GATT negotiating history, developing countries made it clear to the EU and the United States that their differences would not define the outcome of a meeting. The result was a decision by ministers to suspend the Uruguay Round for four months giving time to negotiators to come up with a substantive compromise on the outstanding issues, and making agreement on all other matters conditional on consensus regarding progress on agriculture. The Montreal meeting made clear that the EU had to make substantive commitments to liberalize trade in agriculture. Not doing so would mean the end of the Uruguay Round. The compromise that was hammered out in the subsequent months largely revolved around ideas proposed by the Cairns Group (see the next section).

During 1989 90 negotiations in all areas continued, the goal being to finalize negotiations at a December 1990 ministerial meeting. This meeting, held in Brussels, also ended in failure reflecting substantive disagreements in a number of areas, but particularly in agriculture. The failure did not come as a surprise. The draft Final Act submitted to

ministers had a huge amount of "square-bracketed" text, indicating alternative options or proposals on a subject, as well as "open questions." The unwillingness of the EU to make policy reform commitments on a commodity and instrument basis a key demand of exporting countries forced the round into overtime. The bottleneck was very similar to what would emerge in the Doha Round strong resistance by the EU to suggestions that border protection be reduced and export subsidies constrained.

Talks continued through 1991, leading to a new draft Final Act in December 1991 encompassing some 30 negotiated agreements, including a proposed framework for market access commitments. Again it was agriculture that precluded agreement, with the EU and several in the Cairns Group rejecting the overall package. Although another two years of negotiations would be needed to conclude the deal, in the end the contours established by the 1991 draft final act the so-called Dunkel text defined what was ultimately agreed. What was missing were the specific market access commitments of the countries participating in the negotiations. This became the focal point for bargaining in 1992 93, which eventually led to a deal under the leadership of a new Director-General, Peter Sutherland, appointed in mid-1993.

In June 1993, the U.S. Congress granted an extension of fast-track authority to the U.S. Administration under which Congress could not propose amendments to the outcome of negotiations setting a December 15 deadline for talks to be concluded. This proved to be an effective focal point the substantive negotiations were concluded on December 13, 1993. In Marrakesh, on April 15, 1994, ministers signed the Final Act establishing the World Trade Organization (WTO) and embodying the results of the Uruguay Round.

Key Issue Areas What follows briefly discusses a number of the main negotiating topics that figured in the final "grand bargain."

Tariffs In contrast to the previous two rounds (Kennedy and Tokyo), in the Uruguay Round negotiators did not use a formula approach to cut tariffs. Instead they reverted to an item-by-item, request-offer approach. This reflected the U.S. view that because average MFN tariffs were low, it made more sense to focus talks on specific sectors and on tariff peaks. Although a large number of countries would have preferred to use a formula approach, the United States prevailed. The result of the negotiations was to lower the average bound tariffs on manufactured products of industrial countries, weighted by the volume of trade in the products concerned, from 6.4 percent to 4.0 percent, a cut of some 40 percent. This compares with a weighted-average duty of about 25 percent before the creation of GATT (1947) and around 15 percent at the time of the Dillon Round (the early 1960s). Developing-country participation as measured by the scope of tariff bindings increased substantially during the Uruguay Round, with the percentage of bound lines increasing from 22 to 72 percent. In addition, all countries agreed to bind 100 percent of their agricultural tariff lines.

Agriculture The Punta del Este declaration broke new ground in making an explicit reference to liberalization, with all policies affecting agricultural trade to be discussed, including domestic and export subsidies. This contrasted with the Kennedy and Tokyo Round ministerial declarations, which emphasized the status of agriculture as a special (unique) sector and were oriented toward the negotiation of commodity-specific agreements. The Uruguay Round talks on agriculture were largely between the United States and a group of 14 other agricultural exporters that sought significant liberalization, and a set of countries that provided heavy protection for their farmers: the EU, other European countries (Switzerland, Norway, and Finland), Japan, and South Korea.

The coalition of 14 exporters was called the Cairns Group, after the Australian city that hosted the meeting at which it was established. Members included Argentina, Australia, Brazil, Canada, Chile, Colombia, Fiji, Hungary, Indonesia, Malaysia, New Zealand, the Philippines, Thailand, and Uruguay. It therefore had both high-income and developing country members, illustrating that on agriculture in contrast to IPRs, safeguards, and textiles and clothing the negotiations did not involve a North-

South split. The objective of the Cairns Group was to gradually attain free trade in agricultural commodities, eliminate production distortions, and ensure that binding undertakings to this effect were negotiated. The United States initially called for the complete liberalization of trade in agriculture, including the abolition of all export subsidies. The EU, in contrast, initially proposed that negotiations adopt the approach used in past GATT talks: focus on "emergency measures" for specific commodities such as cereals and sugar to reduce supply and seek to stabilize world agricultural markets. The differences between the United States and the EU were therefore very great: one proposing free trade, the other "managed trade." The Cairns Group was less extreme in pushing for liberalization than the United States, but took a position much closer to that of the United States than that of the EU and other protectionist countries.

Bridging the gap between the EU and the U.S.-Cairns positions proved extremely difficult, not only because of fundamental, substantive differences, but also because of the negotiating strategies that were pursued. Although it was clearly unacceptable to the EU, for the first two years the United States maintained its demand for the total elimination of trade-distorting support policies within 10 years. As already noted, the resulting standoff led to the breakdown of the Montreal midterm review of the round in December 1988. The April 1989 compromise was that the objective would be "substantial progressive reductions in agricultural support," to be implemented through specific commitments and/or agreement on a maximum "aggregate measure of support" (AMS).

At the December 1990 ministerial meeting in Brussels, the EU refused to accept the compromise text proposed by the chairman of the negotiating group which would have averaged a cut of about 25 percent in protection levels. This had significant implications for EU farm policy, the reform of which was under active discussion at the time. The EU needed to settle its internal debates on agriculture first in particular to placate the French, who opposed any significant move toward meeting U.S.-Cairns Group demands. Latin American members of

the Cairns Group played a major role in opposing any significant weakening of the chairman's proposed text. Argentina and Brazil refused to accept the proposed deal and were ready to scuttle the Uruguay Round over the issue.

An agreement between the EU and the United States was eventually reached after much brinkmanship with the so-called Blair House Accord in November 1992. Under the deal, EU policies to pay farmers to take land out of production would not be included in the definition of the Aggregate Measure of Support, nor would similar U.S. payments. The two players also agreed that the proposed 20 percent reduction in the AMS would be an average cut and not apply on a product-by-product basis and that the volume of exports benefiting from export subsidies would be cut by 21 percent.

Textiles and Clothing Liberalization of trade in textiles and clothing was a key demand of developing-country exporters, in particular those that were most restricted under the MFA. Negotiations were difficult. Major areas of disagreement concerned the modalities of phasing out the MFA, the implementation period, and the need for special safeguards. All these matters were eventually negotiated without the brinkmanship and breakdowns that characterized the agricultural negotiations. The deal stipulated that the MFA would be phased out over a 10-year period (1995 2004) in four stages. In 1995, quantitative restrictions affecting at least 16 percent of product lines were to be removed, in 1998 another 17 percent, in 2002 a further 18 percent, followed by the remaining 49 percent at the end of 2004. This did not imply the removal of all trade barriers: tariffs would continue to be imposed. As a result of the Uruguay Round, average tariffs were reduced to a trade-weighted average of 12.1 percent, down from 15 percent.

Services The move to consider rules for trade in services originated with the United States, which had a large services trade surplus. Early in the negotiations, many developing countries argued for a narrow definition of trade in services, excluding sales of services by foreign affiliates of multinational firms. These countries also emphasized the need for

governments to remain unconstrained in their freedom to discriminate against foreign suppliers of services, including inward foreign direct investment (FDI). A consequence was outright rejection of any suggestion that the principle of national treatment apply to services. The EU disagreed with developing countries on how trade in services should be defined, arguing that all types of transactions required to achieve effective market access should be covered, including FDI. The EU agreed, however, that a multilateral agreement on trade in services should not necessarily involve far-reaching obligations of a generally binding nature. Specifically it proposed that national treatment be a specific commitment, to be negotiated on a sector-by-sector basis. As was the case in agriculture, the United States made the most far-reaching proposal, arguing that the nondiscrimination principle should apply to all countries as a binding, general obligation and that all measures limiting market access should be on the table.

Given the lukewarm interest of many developing countries in services negotiations, the EU position emerged as the compromise. In return for acceptance that trade in services be defined broadly to include FDI, national treatment would be sector specific and not a general commitment. In addition, commitments could be made on a limited number of specific policies deemed to restrict access to markets. Virtually all commitments made in the Uruguay Round were of a lock-in nature, that is, a promise not to become more restrictive than what was implied by actually prevailing policies for specific sectors.

In the closing days of the Uruguay Round it became clear that it would be difficult to come to closure on a number of services sectors, including financial services, basic telecommunications, maritime transportation, and one important mode of supply: movement of natural persons (service providers). (Air transportation was excluded from the negotiations altogether.) Rather than allow a situation to develop in which countries would withdraw previously negotiated commitments in these areas, it was agreed that talks on these services were to continue after the conclusion of the Uruguay Round. In addition, on a number of subjects where no agreement proved possible   subsidies, procurement, and safeguards   specific rules were left for future deliberations.

IPRs The negotiation on IPRs was one of the more difficult of the Uruguay Round, both politically and technically. The subject was new to the GATT. Moreover, in contrast to agriculture, it involved a North-South confrontation. Industrial countries, led by the United States, sought an ambitious and comprehensive agreement on standards of protection for IPRs, enforceable through the dispute settlement mechanism. The G10 and other developing countries sought to differentiate discussions on trade in counterfeit goods from IPRs more broadly defined, indicating a willingness to cooperate on the former but not the latter. A general concern was that greater protection of IPRs would strengthen the monopoly power of multinational companies and detrimentally affect poor populations by raising the price of medicines as well as other products.

The first two years of negotiations were dominated by disagreements over the mandate of the negotiating group. Areas of disagreement included standards of protection, use of unilateral sanctions, and the need for   and length of   transitional periods. One of the most difficult questions was whether it was acceptable for GATT contracting parties to draft substantive standards for IPRs. Some developing countries, led by India, argued that the trading system should not be in the business of setting or enforcing IPRs. This was a task for the World Intellectual Property Organization and national governments.

At the end of the day the outcome of negotiations was a far-reaching and comprehensive set of disciplines on IPRs. One reason why this was accepted by developing countries was that they obtained concessions in other areas that mattered more to them. Also important was that by agreeing to the inclusion of disciplines on IPRs, industrialized nations would no longer be able to impose unilateral sanctions on countries that they deemed to be violating IPRs. The United States in particular was an active user of such sanctions. Given a multilateral agreement on IPRs,

multilateral dispute settlement procedures would apply. Finally, although the negotiations involved a confrontation between developing countries, led by India, and OECD nations, led by the United States, a number of developing countries felt that stricter IPR protection would be in their interest because it would encourage FDI, technology transfer, and innovation.

Safeguards and VERs Contingent protection was a key issue in the Uruguay Round. In the Tokyo Round, no agreement had proven possible on rules pertaining to safeguard actions   the reimposition of import barriers to protect domestic industries from serious injury caused by import competition. As noted previously, due to the global macroeconomic situation in the late 1970s and early 1980s, an increasing number of countries were imposing protectionist instruments, including in particular VERs. A major objective of developing country exporters   especially in East Asia   was to outlaw the use of these measures and to achieve stronger disciplines on the use of quantitative restrictions. In contrast to the Tokyo Round, these objectives were achieved, with not only agreement to abolish the Multifiber Arrangement that constrained trade in textiles and apparel, but also a new agreement on safeguards that banned VERs.

Other Subjects Space constraints preclude a discussion of the many other areas in which negotiations occurred. These resulted in many new disciplines on matters such as import licensing, customs valuation, product standards, nonagricultural subsidies, state-trading enterprises, TRIMs, transparency, and dispute settlement. The last subject was an important one. As a result of the round, significant changes were made to the way disputes are dealt with. Under the new WTO, countries would no longer be permitted to block adoption of dispute settlement findings by arbitration panels, instead being limited to appealing findings before a new Appellate Body whose rulings would be final.

Developing Countries Developing countries participated actively in the Uruguay Round negotiations and had a significant impact on the outcome. It became obvious that it was no longer appropriate to regard developing countries as a bloc (assuming this had ever been the case). Instead, countries pursued their self-interest and teamed up with high-income countries on a number of issues. The Cairns Group was just the most prominent example of a North-South coalition.

In contrast to the Kennedy and Tokyo Rounds, the Uruguay Round was a single undertaking: all agreements were to apply to all members. As a result, developing countries became subject to a large number of new obligations   some newly negotiated in the Uruguay Round, others originally negotiated during earlier rounds among industrialized nations. In recognition of the fact that the implementation burden associated with the round would fall heavily on developing countries, they were generally given longer transition periods.

The WTO The 1986 Punta del Este ministerial declaration made no mention of the possible creation of a new organization. A 1990 Canadian suggestion to establish a multilateral trade organization   supported by the EU   was therefore something of a surprise. The proposal was motivated by a desire to create a single institutional framework to encompass the modified GATT and the new agreements on services (the GATS) and intellectual property (TRIPS), as well as all other agreements and arrangements concluded under the auspices of the Uruguay Round. The United States initially opposed the idea of a new institutional framework, but eventually agreed to the creation of the World Trade Organization.

The Uruguay Round was a watershed event in the history of the trading system. In creating the WTO, it led to the realization of the original vision of those who had sought to create an international trade organization 50 years earlier but failed. It involved very active participation by developing countries and pragmatic cooperation between rich and poor nations on specific issues of mutual interest, and it illustrated that very politically sensitive policies could be subjected to multilateral disciplines. The "grand bargain" that was negotiated was a very complex one. Concerns that it was not sufficiently balanced became prominent in the

years immediately following the conclusion of the round and were a major factor in the launch of the first round under WTO auspices, the Doha Round.

See also agricultural trade negotiations; Doha Round; General Agreement on Tariffs and Trade (GATT); General Agreement on Trade in Services (GATS); multilateral trade negotiations; World Trade Organization

FURTHER READING

Bhagwati, Jagdish, and Mathias Hirsch, eds. 1998. *The Uruguay Round and Beyond: Essays in Honor of Arthur Dunkel.* Ann Arbor: University of Michigan Press. A compilation of retrospectives by negotiators who participated in the round.

Croome, John. 1999. *Reshaping the World Trading System.* Leiden: Kluwer International. The best negotiating history of the round, discussing in depth the positions taken, compromises made, and the chronology of the negotiations.

Hoekman, Bernard, and Michel Kostecki. 2001. *The Political Economy of the World Trading System.* Oxford: Oxford University Press. Summarizes the law and economics of the Uruguay Round agreements, their genesis, and implementation.

Martin, Will, and L. Alan Winters, eds. 1996. *The Uruguay Round and the Developing Countries.* Cambridge: Cambridge University Press. Provides empirical assessments of the outcome of the negotiations.

Schott, Jeffrey, and Johanna Buurman. 1994. *The Uruguay Round: An Assessment.* Washington, DC: Institute for International Economics. Briefly summarizes and assesses all the results of the negotiations.

World Trade Organization. 2007. *World Trade Report 2007: Six Decades of Multilateral Trade Cooperation.* Geneva: WTO. An official review of the multilateral trading system that attempts to take a "long view" of the subject.

BERNARD HOEKMAN

V

▪ vehicle currency

A vehicle currency is a reserve currency that central banks or monetary authorities use for the purpose of exchange intervention and the pricing of goods, services, and assets entering international trade and finance. A vehicle currency allows the nation issuing the currency to pay a somewhat lower price for imports, borrow at marginally lower rates, and earn seigniorage. The U.S. dollar is by far the most important vehicle currency in the world today. The euro (the currency of the European Monetary Union, or EMU) has also been used as an increasingly important vehicle currency since its creation in 1999.

A vehicle currency fulfills in the world economy the same basic functions that the currency performs in the economy of the nation. It serves as a unit of account (i.e., it is the primary invoicing currency for international commodity and asset trade), as a medium of international exchange (in that global foreign exchange transactions are overwhelmingly conducted in that currency), and as a store of value (i.e., it is the currency in which official reserve assets are primarily held by nations' central banks or monetary authorities). A national currency becomes a vehicle or international currency as a result of market forces that allow it to function as money for both private and official transactions in the world economy. The U.S. dollar serves as the vehicle currency par excellence in global foreign markets. More than 60 percent of the U.S. currency is actually held abroad.

During the 19th and early 20th centuries, the British pound sterling was the dominant vehicle currency. Since then, the international use of the U.S. dollar increased in step with the increase in the relative economic and political importance of the United States. After World War II, the dollar became the dominant vehicle currency. The reasons for the decline of the pound sterling and rise of the U.S. dollar as a vehicle currency after World War II were (1) the high rate of inflation in the United Kingdom and sharp fluctuation in the value of the pound compared with the low inflation in the United States and stability of the U.S. dollar during the late 1940s and early 1950s, (2) the existence of exchange controls in England in contrast to the relative openness of the U.S. financial market, and (3) the decline in the sterling area's share of world exports in comparison with the rise in the U.S. share. Today the pound sterling remains a vehicle currency, although to a much smaller extent than the U.S. dollar, because London remains a sophisticated international financial center. One indication of the changed international roles of the dollar and pound after World War II was the decision by the Organization of the Petroleum Exporting Countries (OPEC) in the mid-1970s to price petroleum in dollars instead of pounds.

Table 1 shows the relative importance of the dollar, the euro, and other major currencies in the world economy in 2004. The table shows that 44.1 percent of foreign exchange trading was in dollars, as compared with 18.6 percent in euros, 10.2 percent in Japanese yen, and smaller percentages in other currencies. The table also shows that 50.3 percent of international bank loans and 48.4 percent of

Table 1
Relative international importance of major currencies in 2004 (percent)

	Foreign exchange trading[a]	International bank loans[a]	International bond offering[a]	Trade invoicing[b]	Foreign exchange reserves[c]
U.S. dollar	44.1	50.3	48.4	52.0	65.9
Euro	18.6	51.0	44.3	24.8	24.9
Japanese yen	10.2	7.6	1.2	4.7	3.9
Pound sterling	8.5	4.3	5.2	5.4	3.3
Swiss franc	3.1	0.4	0.2	na	0.2
Other currencies	15.5	1.6	1.1	13.1	1.8

Sources:
[a]Bank of International Settlements, *Triennial Central Bank Survey* (Basle: BIS, March 2005) and BIS data set.
[b]P. Bekx, "The Implication of the Introduction of the Euro for Non EU Countries," *Euro*, Paper No. 26, July 1998. Data are for 1995.
[c]*IMF, Annual Report* (Washington, DC: IMF, 2005).

international bond offerings were in dollars, and 52 percent of international trade invoicing was denominated in dollars. Also, 65.9 percent of foreign exchange reserves were held in U.S. dollars, as compared with much smaller percentages for the euro and other currencies.

Besides being the most important vehicle currency, the U.S. dollar is also used as the currency in a number of Latin American countries (Ecuador, El Salvador, Guatemala, and Panama), as well as in the Commonwealth of Puerto Rico and the U.S. Virgin Islands in the Caribbean, in a process called dollarization. Many developing countries also peg their currency to the U.S. dollar. Similarly, a number of small states use the euro as their national currency. Besides the 12 countries of the Eurozone, the euro also is the legal tender of Monaco, San Marino, Vatican City, Andorra, Montenegro, and Kosovo, as well as of the Eurozone overseas territories of French Guiana, Guadeloupe, Martinique, Mayotte, Reunion, and Saint Pierre et Miquelon. Slovenia adopted the euro on January 1, 2007. Cyprus, Estonia, Latvia, and Malta were scheduled to adopt the euro in 2008; Bulgaria, Slovakia, and Lithuania in 2009; the Czech Republic and Hungary in 2010; and Poland and Romania in 2011.

In recent years, the United States has faced unsustainably large trade deficits (exceeding 6.5 percent of gross domestic product GDP in 2005). Reducing U.S. trade deficits to sustainable levels (say, to 2 or 3 percent of GDP) may require, among other things, a large depreciation of the dollar. This could lead central banks or monetary authorities (particularly Asian ones, which are by far the largest foreign holders of dollar reserves) to reduce their holding of dollars and induce petroleum-exporting countries to switch to euros in the pricing of petroleum, and thus put at risk the vehicle-currency status of the dollar. Similar risks for the dollar arose at the time of the collapse of the Bretton Woods system in 1971, in the face of large U.S. trade deficits and the U.S. suspension of the dollar convertibility into gold. The dollar survived, however, and even increased its vehicle-currency status during the 1980s and 1990s because no other currency could assume that role. The situation is different today because of the existence of the euro as a possible alternative to the dollar. Although the vehicle-currency status of the dollar seems secure for now, the importance of the euro (and eventually of the Chinese yuan) is likely to grow at the expense of the dollar in the future (especially if Britain also adopts the euro), and these currencies could become as important vehicle currencies as the dollar if U.S. trade deficits remain excessive and the international value of the dollar falls significantly.

See also Bretton Woods system; convertibility; currency substitution and dollarization; dollar standard; dominant currency; euro; exchange rate regimes; foreign exchange intervention; global imbalances; gold standard, international; International Monetary Fund (IMF); international reserves; money supply; petrodollars, recycling of; reserve currency; seigniorage; special drawing rights; twin deficits

FURTHER READING

Bank of International Settlements (BIS). 2006. *Annual Report*. Basel: BIS. Report by the Bank for International Settlements on the world's financial and banking situa tion, including the use of vehicle currencies.

Cohen, B. J. 2000. *Life at the Top: International Currencies in the Twenty First Century*. Princeton Essay in Interna tional Economics No. 221 (December). Princeton, NJ: Princeton University. An excellent discussion of the ve hicle currencies during this century.

Golberg, Linda, and Cedric Tille. 2005. "Vehicle Currency Use in International Trade." Federal Reserve Bank of New York Staff Report No. 2000 (January). Estimates of the relative use of various vehicle currencies in interna tional trade invoicing.

International Monetary Fund (IMF). 2006. *Annual Report*. Washington, DC: IMF. An analysis of the world's eco nomic and financial situation by the International Monetary Fund; it includes a discussion of the relative use of various vehicle currencies.

Krugman, Paul. 1992. "The International Role of the Dollar." In *Currencies and Crises*, edited by Paul Krug man. Cambridge, MA: MIT Press, chapter 10. An analysis of the international role of the dollar before the creation of the euro.

Mundell, Robert A. 2000. "A Reconsideration of the Twentieth Century." *American Economic Review* 90 (June): 327 40. A review and interpretation of the op eration of the international monetary system during the past century by a Nobel Prize recipient in economics.

"The Passing of the Buck?" 2004. *The Economist* (December 2), 78 82. A discussion of the possibility of the dollar losing its primary vehicle currency role to the euro.

Salvatore, Dominick. 2004. "Euroization, Dollarization, and the International Monetary System." In *Monetary Unions and Hard Pegs*, edited by Volbert Alexander et al. New York: Oxford University Press, 27 40. Discusses the use of the euro and the dollar as vehicle currencies and their relationship to the operation of the present international monetary system.

Tavlas, George. 1998. "The International Use of Curren cies: The U.S. Dollar and the Euro." *IMF Finance and Development* 35 (June): 46 49. Discusses trends in the international use of the dollar and assesses the possible impact of the euro in world financial markets.

DOMINICK SALVATORE

■ vertical versus horizontal foreign direct investment

Horizontal foreign direct investment (FDI), in which multiplant firms duplicate roughly the same activities in multiple countries, has been distinguished from vertical FDI, in which firms locate different stages of production in different countries. The bulk of FDI is horizontal rather than vertical. That developed countries are both the source and the host of most FDI suggests that market access is more important than reducing production costs as a motive for FDI.

Brainard (1993) reports that foreign affiliates owned by U.S. multinationals export only 13 percent of their overseas production back to the United States, so most production by U.S. multinationals appears to be motivated by the desire to serve markets abroad. Similarly, the U.S. affiliates of foreign multinationals export only 2 to 8 percent of their U.S. production back to their parents; 64 percent is sold in the U.S. market. The bulk of FDI is attracted to big markets, rather than to cheap labor (or other factors of production). The large volume of two-way FDI flows also seems to fit horizontal FDI models better than vertical ones.

Standard models of horizontal FDI revolve around the trade-off between plant-level fixed costs and trade costs (see Markusen 1984). When the potential host country is small, the potential savings in trade costs (which accrue per unit of exports to the country) are insufficient to offset the fixed costs of setting up a production facility there; hence, exports are chosen over FDI as the method for serving the

market abroad. When a host country is large enough for the fixed costs of the plant to be offset by the trade costs saved, however, FDI is chosen over exports. Bigger market size of the host country, smaller plant-level fixed costs (smaller plant-level scale economies), and larger trade costs are more conducive to horizontal FDI. The *proximity-concentration hypothesis* refers to the common tenet that FDI occurs when the benefits of producing in a foreign market outweigh the loss of scale economies that could be reaped if produced in only one plant (in the firm's home country).

FDI may exist to avoid not only actual trade costs but feared trade costs as well. FDI, such as by Japanese firms into the European Union (EU) in electronics and into the Unites States in autos, may be motivated more by fear of impending trade barriers (anti-dumping duties or voluntary export restraints) than by any barriers in place at the time of the investments.

When the choice between FDI and exports involves a simple trade-off between trade costs and fixed costs, an interesting implication is that no firm should simultaneously engage in both FDI and exports. Even for the exact value of trade costs where the trade costs times the number of units exported equals the plant-level fixed costs, when the firm is exactly indifferent between FDI or exports, the firm will either pay the fixed costs to build the plant and serve the market exclusively through FDI, or not build the plant and serve the market only through exports.

Unlike horizontal FDI, with vertical FDI, firms engage in both FDI and exports. Whereas in horizontal FDI models, the two countries are often envisioned as being of similar size, in vertical FDI models, the home country is usually thought of as being much larger than the host country. Thus the horizontal FDI framework is more representative of a pair of developed countries, whereas the vertical FDI framework is like a developed source country and a developing host country. In horizontal FDI models, the question is how best to serve the host market (abroad), whereas in vertical FDI models, the question is typically how best to serve the domestic market.

Standard models of vertical FDI involve deciding where to locate production to minimize costs. Headquarters services are located in the home country; production of the good can be located with the headquarters in the home country or else separated from headquarters and located abroad. Production costs are assumed to be lower in the host country than at home. Hence the trade-off is between the lower costs of producing abroad and the need to pay trade costs to bring the goods back home. FDI occurs if the cost savings from producing abroad are greater than the trade costs incurred. Lower trade costs should encourage vertical FDI but discourage horizontal FDI. As trade costs fall, vertical FDI occurs for smaller differences in factor prices. In a simple setup where only one unit of labor is required to produce the good in either country, vertical FDI occurs if the wage difference across countries is greater than the trade costs. As vertical FDI is often called international *outsourcing/offshoring*, the production cost savings minus the trade costs can be called the gain from offshoring.

Anyone fearing that, as trade costs fall, all production will shift from rich countries such as the United States to poorer countries such as China or Mexico (where wages are lower) should bear in mind that the United States remains the largest recipient of FDI inflows. Also the comparison is not of wage levels alone, but efficiency wages labor costs per unit of production. If wages elsewhere are one-tenth U.S. wages but workers are less than one-tenth as productive, labor there is not truly less expensive.

The *knowledge-capital model of the multinational enterprise* is an overarching model that includes both horizontal and vertical FDI as special cases. It has been used to test for evidence in support of horizontal versus vertical FDI. Most findings have been more supportive of horizontal FDI, but other research (such as Braconier, Norback, and Urban 2005) has emphasized that vertical FDI does indeed occur and is important to the host countries in which it occurs (sales by affiliates are large relative to GDP). Both

horizontal and vertical FDI can occur in Markusen and Venables (2005)   the split between market-oriented and export platform activity depends mostly on trade costs, and factor endowments influence whether to specialize in components or assembly.

Given that the bulk of FDI is horizontal in nature, and that horizontal FDI is motivated by avoiding trade costs (tariff jumping), the trends in the 1990s were rather perplexing. Dramatic reductions in trade costs due to trade negotiations and technological change occurred together with substantial growth in FDI (outpacing the fast growth in world trade). Neary (2008) has put forth two potential explanations. He shows that cross-border mergers can be encouraged by reductions in trade costs. As mergers and acquisitions are quantitatively more important than greenfield investments (building from scratch), falling trade costs can be consistent with expansions in horizontal FDI. He also argues that horizontal FDI in trading blocks can be encouraged by trade liberalization within the trade block. When trade costs fall within the block, outside firms invest in one country as a means for serving the entire trade block. For example, a U.S. firm may produce in Ireland to serve all of Europe, or a German firm may produce in Canada to serve all of North America. These are examples of export-platform FDI, discussed below. Which explanation is most empirically relevant remains to be determined.

Export-Platform FDI Export-platform FDI is FDI motivated by a desire to export rather than to serve the local market. Vertical FDI is export-platform FDI where the exports are sent back to the home market. However, there is an increasing trend toward export-platform FDI where the exports are sent to third markets. The rise of trade blocks with low internal trade barriers but higher external barriers may contribute to this trend. Multinationals are establishing production subsidiaries within a trade block and using that plant to serve the entire block. To the degree that the host country is small relative to the overall size of the trade block, the vast bulk of production will be exported to other countries in the trade block.

Motta and Norman (1996) find that improved market access within a trade block leads to export-platform FDI in this manner. As an additional benefit, since FDI into the block becomes more attractive to outside firms, due to firms being better able to reach the majority of markets within the block through exports from one plant, the subsidies required to entice firms to locate in the block will be reduced. Instead of considering only the market size of a potential host country, firms now consider the broader, regional market that can be easily reached from the country. As trade blocks are often formed on a regional basis, avoiding artificial trade barriers (such as tariffs) and natural trade barriers (transportation costs) tend to go hand in hand.

Kumar (1998) emphasized the need to distinguish between export-platform FDI oriented toward the home market versus that oriented toward third countries. FDI for export back to the home market occurs to take advantage of cheaper factors of production elsewhere, and only trade costs between the home and host country matter. FDI for export to third countries, in contrast, is critically dependent on the ease of access to the third countries, and the trade costs back to the home market matter little.

Ekholm, Forslid, and Markusen (2007) further distinguish between three types of export-platform FDI. Home country export-platform FDI involves export back to the parent. Third-country export-platform FDI involves export to another large country (not home or host). With global export-platform FDI, the host plant exports to both the home country and the third country. When the home and the host countries form a free trade area, the outcome can be that the inside firm engages in home (or global) export-platform FDI, while the outsider firm opts for the third-country approach. Fitting this scenario, the North American affiliates of U.S. multinationals concentrate on exports back home, whereas affiliates in Europe concentrate on exports to third countries (see also Yeaple 2003). With the North American Free Trade Agreement, Mexico has seen increases in the share of production

by affiliates of multinationals (both U.S. and from elsewhere) sent to the United States.

Using data for U.S. outbound FDI to members of the Organisation for Economic Co-operation and Development from 1980 to 2000, Blonigen et al. (2004) find evidence consistent with export-platform FDI in Europe. When measures of market potential (size of proximate third country markets) are included, they find a clear negative relationship between FDI into proximate countries. This pattern of substitution between industrialized countries in Europe provides strong evidence of export-platform FDI. Ireland is the EU economy most successful in attracting export-platform FDI.

The implications of export-platform FDI need further study. For example, when multinational enterprises (MNEs) use the host country as an export platform, local firms are often not competitors (unless also exporting) and thus the MNEs need not worry about restricting technology spillovers. As there is less risk of damaging local competitors, local governments may view export-platform FDI more favorably than FDI for the purpose of serving the local market. While there are potential employment gains from both, export-platform FDI does not generate the gains in consumer surplus that market-access motivated FDI would.

See also fixed costs and foreign direct investment; knowledge-capital model of the multinational enterprise; market size and foreign direct investment; outsourcing/offshoring; proximity-concentration hypothesis; trade costs and foreign direct investment

FURTHER READING

Blonigen, Bruce A., Ronald B. Davies, Glen R. Waddell, and Helen T. Naughton. 2004. "FDI in Space: Spatial Autoregressive Relationships in Foreign Direct Investment." NBER Working Paper No. 10939. Cambridge, MA: National Bureau of Economic Research. Available at http://www.nber.org/papers/w10939.pdf. Finds evidence consistent with export platform FDI in Europe.

Braconier, Henrik, Pehr Johan Norback, and Dieter Urban. 2005. "Multinational Enterprises and Wage Costs: Vertical FDI Revisited." *Journal of International Economics* 67 (2): 446 70. Evidence in support of vertical FDI.

Brainard, S. Lael. 1993. "An Empirical Assessment of the Factor Proportions Explanation of Multinational Sales." NBER Working Paper No. 4583. Cambridge, MA: National Bureau of Economic Research. Available at http://nber15.nber.org/papers/w4583.pdf. Finds little support for factor proportion motivation for FDI.

Ekholm, Karolina, Rikard Forslid, and James R. Markusen. 2007. "Export Platform Foreign Direct Investment." *Journal of the European Economic Association* 5 (4): 776 95. Examines conditions under which three different types of export platform FDI arise: back to MNE's home country, to a third country, or globally (to both).

Kumar, Nagesh. 1998. "Multinational Enterprises, Regional Economic Integration, and Export Platform Production in the Host Countries: An Empirical Analysis for the U.S. and Japanese Corporations." *Weltwirtschaftliches Archiv/Review of World Economics* 134 (3): 450 83. Contrasts the determinants of production for export back to MNE's home market to those for export to third country markets.

Markusen, James R. 1995. "The Boundaries of Multinational Enterprises and the Theory of International Trade." *Journal of Economic Perspectives* 9 (2): 169 89. Excellent review of empirical evidence on multinational firms.

———. 1984. "Multinationals, Multiplant Economies, and the Gains from Trade." *Journal of International Economics* 16 (3 4): 205 26. The original model of horizontal FDI.

Markusen, James R., and Anthony J. Venables. 2005. "A Multi Country Approach to Factor Proportions Trade and Trade Costs." NBER Working Papers No. 11051. Cambridge, MA: National Bureau of Economic Research. Available at http://www.nber.org/papers/w11051.pdf. Model where both horizontal and vertical FDI occur.

Motta, Massimo, and George Norman. 1996. "Does Economic Integration Cause Foreign Direct Investment?" *International Economic Review* 37 (4): 757 83. Three country, three firm model showing how improved market access within a region leads to export platform FDI as outside firms invest in the region.

Neary, J. Peter. 2008. "Trade Costs and Foreign Direct Investment." In *Foreign Direct Investment and the Multinational Enterprise*, edited by Steven Brakman and Harry Garretsen. Cambridge, MA: MIT Press, 13 38. Provides two explanations for how falling trade costs can generate increased FDI, as seen in the data.

Yeaple, Stephen Ross. 2003. "The Complex Integration Strategies of Multinationals and Cross Country Dependencies in the Structure of Foreign Direct Investment." *Journal of International Economics* 60 (2): 293 314. Describes how strategies of multinationals are richer than just horizontal or just vertical FDI.

AMY JOCELYN GLASS

■ **voluntary export restraints**
See quotas

■ **wages**

See trade and wages

■ **Washington consensus**

The term *Washington consensus* was coined by John Williamson (1990) to encapsulate the set of policy reforms advocated with a reasonable degree of consensus by international financial institutions, the U.S. government, the Federal Reserve Board, and the leading think tanks based in Washington. Those policies were deemed necessary to achieve growth, low inflation, a viable balance of payments, and equitable income distribution in the developing world at large, and especially in Latin America, which was still recovering from the debt crisis that erupted in 1982. The policies that defined the Washington consensus included (1) fiscal discipline, (2) increased public expenditure on social services and infrastructure, (3) tax reform to broaden tax bases and reduce marginal tax rates, (4) market-determined interest rates, (5) unified and competitive exchange rates, (6) import liberalization, (7) openness to foreign direct investment, (8) privatization, (9) deregulation, and (10) secure property rights.

As the term evolved in the 1990s, it became synonymous with the main policies promoted consensually by the International Monetary Fund (IMF), the World Bank, and the U.S. Treasury, with an emphasis on macroeconomic stability through sound fiscal and monetary policies, and a wider role for markets in relation to governments in the allocation

of resources through privatization, trade and capital account liberalization, and domestic market deregulation. The prescription to stabilize, privatize, and liberalize had gained traction among many policymakers in developing countries since the 1980s as import substitution and interventionist policies had led to bloated bureaucracies, grossly inefficient state-owned enterprises, unsustainable public indebtedness, inflation, and slow or even negative growth. The collapse of the Soviet bloc and the success of countries, such as Chile, that had already experimented with the new policies, contributed to their acceptance in the early 1990s. The limited role of government and the expanded role of markets have led some to refer to Washington consensus policies as "neoliberal" or based on "market fundamentalism," an expression that does not do justice to the scope and richness of the policies, especially those originally listed by Williamson (1990). This entry adopts the more recent and more widely used meaning of Washington consensus policies, in spite of its partial overlap with Williamson's list.

The Set of Policies For convenience, Washington consensus policies can be divided into macroeconomic and microeconomic, or structural, policies. The macroeconomic policies are mostly concerned with achieving price stability (more strictly, low inflation rates) and preventing foreign exchange or public debt crises. The cornerstone of macroeconomic stability is fiscal discipline. This need not imply a balanced budget, but rather a fiscal position that does not lead to a rising debt to gross domestic product (GDP) ratio, loss of international

currency reserves, or (unacceptably high or rising) inflation (caused by monetary financing of the deficit), and therefore is "sustainable" without major policy adjustments.

Monetary discipline, a second key element of macroeconomic stability, is best achieved when monetary policy instruments (the intervention interest rate, open market operations, etc.) are aimed at controlling the supply of money and/or the price level and not at reaching other objectives such as accelerating growth or reducing unemployment. Central bank independence from government and other sources of political interference is additionally seen as facilitating monetary discipline.

Though no consensus exists in regard to the exchange rate regime most conducive to macroeconomic stability, it is well established that fixed or pegged exchange rate systems make monetary policy instruments ineffective in an environment of international capital mobility. Furthermore, fixed exchange rates insufficiently backed by international reserves may induce speculative attacks that often lead to abrupt exchange rate corrections. Nonetheless, fully floating exchange rates are seldom desirable, as central banks cannot allow exchange rates to freely fluctuate when a large portion of government or private sector debts are denominated in foreign currency and therefore exposed to exchange rate risks, as is the case in most Latin American countries (an attitude known as "fear of floating"). Thus international financial institutions often advocate for strong international reserve positions and limited reliance on external finance as precautionary measures to weather potential external shocks.

Given a sound macroeconomic framework, structural policies are primarily intended to stimulate and sustain growth by facilitating the functioning of markets and minimizing government interference in economic agents' decisions regarding investment, saving, consumption, and work. This entails lifting tariff barriers and other restrictions on international trade (*trade liberalization*); eliminating restrictions on foreign direct investment and on the free movement of funds in and out of the country (*capital account liberalization*); privatizing state-owned en-

terprises and opening to private investment industries previously controlled by the public sector, such as telecommunications, electricity, or mining (*privatization*); reducing bank reserve requirements and dismantling interest rate controls, subsidized loan programs, and directed credit systems (*financial liberalization*); simplifying the tax code, lowering tax rates, and widening tax bases (*tax reform*); facilitating labor hiring and firing, eliminating wage controls, simplifying the labor code, and reducing payroll taxes (*labor deregulation*); and lifting price controls and license requirements, and easing the procedures for entry and exit of firms (*goods markets deregulation*).

The Extent of Policy Reform Little reform had taken place in the developing world before 1989, but the launch of the Brady Plan by the U.S. government (an initiative to convert the nonperforming bank debts of the developing countries into long-term bonds to restore access by those countries to international finance) and the fall of the Berlin Wall that year signaled the beginning of ambitious macroeconomic and structural reforms in Latin America, Eastern Europe, and the former Soviet Union.

The extent of macroeconomic reforms in the following decade and their degree of success were remarkable. The incidence of high inflation rates among developing countries declined sharply, and median inflation rates among medium-income countries declined from 16 percent in 1990 to 6 percent a decade later. The overall fiscal deficit of developing countries, which had reached 7 percent of GDP in the early 1980s, has remained approximately 2 3 percent of GDP since 1990. By the end of the 1990s, the median developing country posted a primary surplus (the fiscal balance excluding interest payments). In addition, the volatility of key macroeconomic variables such as economic growth and real exchange rates declined in the 1990s, though partially in response to a more stable external environment.

The extent of microeconomic, or structural, reform since the late 1980s has also been substantial, especially in Latin America and the countries of the former Soviet bloc, although the effects of this reform remain a matter of intense debate. Aside from Chile,

which slashed import tariffs and lifted other restrictions on international trade, liberalized the financial sector, privatized many state-owned enterprises, and streamlined the tax and labor codes under the Pinochet dictatorship (1973 90), no other country in Latin America had fully embraced structural reform until the late 1980s, when the Brady Plan created the opportunity of attracting much-needed external capital by adopting the "right" set of policies in the view of international financial institutions and investors. According to reform indexes computed for Latin America, trade and financial liberalization took off first and advanced furthest in the majority of countries. Privatization and tax reforms were more uneven in depth and timing, while labor reform was timid and erratic. Argentina, Bolivia, and Peru were among the most aggressive reformers, while Costa Rica, Ecuador, Mexico, Uruguay, and Venezuela undertook more cautious but nonetheless substantial reforms. Reform lost momentum in the second half of the 1990s, and major reversals took place after 2000 in Argentina, Bolivia, and Venezuela as governments reintroduced a variety of price controls and international trade restrictions. In some cases entire sectors were expropriated, as in the case of Bolivia's natural gas sector.

The reform process was no less turbulent in the former Soviet bloc countries, where it was part of the larger project of state formation that followed the breakup of the USSR (making it more difficult to judge the success of the new policies). Wage and price deregulation, two of the most important components of labor and goods markets deregulation, were undertaken in most countries at the beginning of the transition process, but then stalled and even reversed in some countries after the mid-1990s. A dozen years after the collapse of the Soviet bloc, domestic markets were still almost as heavily regulated as before in Belarus, Moldova, and Uzbekistan. The liberalization of external trade, investment, and finance, on the other hand, has represented a more gradual but sustained process, although as of the first years of the 21st century that process remained incipient in Azerbaijan, Belarus, Turkmenistan, and Uzbekistan. Although the transition from communist to market

economies offered the former Soviet bloc countries ample opportunities for privatization, the reform indexes available suggest that only a handful of countries (Armenia, Bulgaria, the Czech Republic, Hungary, Macedonia, and Slovakia) have privatized the vast majority of their economic activity. As of 2001, the process had barely progressed in Slovenia and Turkmenistan and had suffered drastic reversals in Russia and Uzbekistan.

The process of microeconomic reform varied even more across time, country, and area of reform among the African countries. In the area of financial liberalization, Africa went from a completely dirigiste stance in the mid-1980s to freely moving interest rates and greater flexibility in bank operations in the late 1990s. State-owned financial institutions still intermediated a sizable portion of domestic credit, however, and restrictions on the allocation of credit abounded. Trade liberalization also advanced substantially, but many countries continued to impose high import tariffs and import and export licensing restrictions in the late 1990s. African countries were likewise reluctant to engage in capital account liberalization during the 1980s. Around the mid-1990s most countries undertook at least partial reforms, but many countries continue to require some form of authorization for the acquisition of firms by foreign investors and for portfolio investments by foreigners. Countries in Africa privatized less and more reluctantly than Latin American or transition countries. Less than 40 percent of Africa's state-owned enterprises were affected by privatization, and only four countries Ghana, Côte d'Ivoire, Nigeria, and Zambia accounted for a third of all transactions between 1991 and 2001, worth just $9.1 billion.

Assessment Washington consensus policies have been widely criticized for their ineffectiveness in delivering more macroeconomic stability and higher growth their two main objectives as well as for their social and distributional effects.

Although macroeconomic stability vastly improved in the developing world in the 1990s, it failed to reach developed-country levels, and countries remained vulnerable to extreme macroeconomic events such as banking crises, currency crashes, and

sudden interruptions in capital flows, or "sudden stops." Argentina, a close follower of the Washington consensus recommendations, nonetheless experienced a severe economic crisis in 2001 and 2002: the exchange rate was devalued by more than 300 percent, the government defaulted on its debt obligations, and economic activity and employment collapsed. The inadequate policy conditions imposed by the IMF on the countries affected by the Asian crisis of 1998 are also offered as evidence of the failure of the Washington consensus to prevent macroeconomic instability.

Among the main shortcomings of the policies advocated are sequencing problems (particularly the premature liberalization of capital flows without adequate financial regulation), excessive reliance on high interest rates to contain aggregate demand, price and exchange rate pressures (at the risk of worsening or precipitating a financial or a debt crisis), and inadequate fiscal adjustment that exacerbates busts and weakens economic recovery. The Washington consensus is further criticized for inadequate international financial architecture to respond to the needs of countries in crisis or facing speculative attacks and to prevent "sudden stops."

All of these objections, however, are subject to intense debate. There is no widespread agreement on whether capital controls are effective in moderating or stabilizing capital flows, or whether financial systems isolated from international competition can be adequately regulated. When fiscal sustainability is at risk (for instance, as a result of a sudden interruption in credit access or a decline in the country's terms of trade), interest rate increases or public expenditure cuts may be necessary in order to prevent further capital outflows, currency depreciation, and inflation. Finally, there is intense debate regarding the role the IMF and other international financial institutions should play in preventing and responding to crises, as bailouts and other forms of support may encourage lack of macroeconomic discipline.

The effects of Washington consensus policies on growth have been a contentious issue since their outset. The lackluster performance of Latin America in relation to other regions, especially the fast-growing countries of East Asia, is often offered as evidence of their modest effect. Econometric evidence for Latin America and the transition economies of the former Soviet bloc suggests that Washington consensus policies had positive, but only weak and transitory, effects on growth. The main explanations for this modest growth effect include insufficient depth and completeness of policies and reforms, excessive use of macroeconomic policies to achieve price stability rather than promote growth, sequencing problems (capital account liberalization, domestic financial liberalization, or privatization prior to the regulatory reforms needed to prevent financial instability and market manipulation), and a host of other implementation problems, due especially to lack of institutional support to administer trade reforms, facilitate resource reallocation, prevent corruption, and implement regulatory reforms. Other critics contend that stimulating and sustaining growth requires removing only the most binding constraints, rather than completing the ambitious list of reforms prescribed by the Washington consensus or its augmented versions, which incorporate a variety of institutional or "second-generation" reforms as well as social policies.

The view that Washington consensus policies have deleterious social effects is presently so widespread as to seem beyond discussion. But the empirical evidence is mixed at best. In many countries in Latin America and Africa, trade liberalization had a negative effect on employment in the sectors affected by import competition, which was not compensated by employment creation in the exporting sectors. Nonetheless trade liberalization and privatization produced both winners and losers among the poor, depending on their sectors of activity and their consumption baskets, and there is virtually no empirical support for the view that aggregate poverty was increased by these reforms.

Much more debate attends the distributional consequences of Washington consensus policies. The fact that trade openness may increase inequality in developing countries seems to go against standard economic theory (or at least theory rooted in the simplest version of the Heckscher-Ohlin model of

international trade), since trade should increase the income accruing to the relatively abundant factor, presumably unskilled labor. This effect may be muted or even reversed, however, where trade protection before liberalization was focused on labor-intensive products, as in the case of Mexico, or where the most abundant factor of production is natural resources, as could be the case in Russia or some of the South American countries. Contrary to popular belief, privatizing utilities did not hurt the poor, as the quality and coverage of services improved, while the impact of privatization on income distribution was moderate in Latin American countries, but substantial in some of the ex-communist countries.

An Evolving Concept The terms *Washington confusion* and *Washington contentious*, among others, have been used to convey the state of disarray of the economic, institutional, and social policies advocated by international financial institutions more than a decade after the original Washington consensus. Among the international financial organizations and some Washington-based think tanks, the view prevails that the Washington consensus can deliver growth, stability, and equality if it is "augmented" with a variety of institutional reforms to strengthen state capabilities and with more encompassing and ambitious social policies. In academia, however, as well as in some segments of the World Bank and other organizations, it is argued that a more comprehensive list of policy requirements can hardly be a useful guide for governments short of political support, technical expertise, and administrative capabilities. Furthermore, although no one disputes the importance of macroeconomic stability, secure property rights, market-oriented incentives, outward orientation, and government provision of basic public goods, it is becoming widely accepted that there are no standard prescriptions for achieving these goals. If a new consensus is emerging it is based on the recognition that economic policy reform is always constrained by institutional and political economy factors, which are even harder to reform.

See also capital controls; economic development; evolution of development thinking; financial liberalization; Heckscher-Ohlin model; import substitution industrialization; International Monetary Fund (IMF); political economy of policy reform; sequencing of financial sector reform; trade and economic development, international; World Bank

FURTHER READING

Birdsall, Nancy, and Augusto de la Torre. 2001. *Washington Contentious. Economic Policies for Social Equity in Latin America.* Washington, DC: Carnegie Endowment for International Peace, Inter American Dialogue. The best example of the expanded Washington consensus pro moted by leading Washington think tanks.

Kuczynski, Pedro Pablo, and John Williamson, eds. 2003. *After the Washington Consensus: Restarting Growth and Reform in Latin America.* Washington, DC: Institute for International Economics. Puts forward a new economic policy agenda "that the region should be pursuing after a decade that was punctuated by crises, achieved disap pointingly slow growth, and saw no improvement in the region's highly skewed income distribution."

Stiglitz, Joseph E. 2002. *Globalization and Its Discontents.* New York: W. W. Norton. This worldwide bestseller should better be titled "The Washington Consensus and Its Discontents." The author is a Nobel Prize laureate who served as economic advisor to the U.S. government and as chief economist of the World Bank during the heyday of the Washington consensus (1993 2000).

Williamson, John, ed. 1990. *Latin American Adjustment: How Much Has Happened.* Washington, D.C.: Institute for International Economics. The book that coined the term and first assessed the extent of policy reform in Latin America and the Caribbean.

World Bank. 2005. *Economic Growth in the 1990s: Learning from a Decade of Reform,* Washington, DC: World Bank. A balanced and comprehensive review of the achieve ments and shortcomings of the Washington consensus.

EDUARDO LORA

■ **West African Economic and Monetary Union**

See Economic Community of West African States (ECOWAS)

■ World Bank

The World Bank (henceforth "the bank") is a multilateral institution devoted to the promotion, worldwide, of sustainable economic development and poverty reduction. The bank was established as the International Bank for Reconstruction and Development (IBRD), in Bretton Woods, New Hampshire, in July 1944, and its Articles of Agreement entered into force in December 1945. The International Monetary Fund was set up at the same time: together, the fund and the bank became the "Bretton Woods institutions." The role of the fund in the Bretton Woods system was to administer the international monetary system, initially in the form of a system of fixed but adjustable exchange rates (House, Vines, and Corden 2008). The initial rationale for the bank, in a world where dollars were in short supply and international capital flows were limited, was the provision of low-interest finance for postwar reconstruction.

The rapidity of the postwar recovery, and decolonization in the 1960s, meant that IBRD activities came to be refocused on development and in particular on infrastructure investment. The IBRD operated, as it still does, by borrowing in its own name and then lending on to borrowing governments for public sector projects at a small margin. However it was soon found that, despite evident needs, it was difficult to justify investments on these semicommercial criteria in many developing economies. As a result, the International Development Agency (IDA) was established in 1960 to lend capital from IBRD profits, augmented by funds donated by "Part I" governments to poor "Part II" countries, on an interest-free basis with long grace periods. IDA is a development agency, not a financial institution. The post-1960 bank is therefore a more complex hybrid agency than the one established in 1944.

This more complex bank does three things.

The Bank as a Multilateral Lending Institution. The bank borrows at AAA rates on international markets and then on-lends the funds to countries. This remains the largest of the bank's activities. IBRD lending, financed from this source, averaged $12.8 billion per annum out of a total lending av-

eraging $21.5 billion per year over fiscal years 2002 7. The bank's actual lending (i.e., the total of its disbursed and outstanding loans) stood at $98 billion in 2007; its outstanding IDA development credits amounted to $102 billion, making a total of $200 billion (World Bank 2006).

The bank is able to borrow funds at rates lower than the London interbank offered rate because countries are reluctant to default on loans from an international institution, because the bank is able to provide additional lending to prevent default arising from inability to pay, and also because it has a buffer of "callable capital" money that member countries have an obligation to contribute if the bank were ever to get into financial difficulties. The bank lends funds to governments with a margin over the cost-of-funds, which makes a contribution to the bank's operational costs. This margin includes a commitment fee and has typically been in the region of 25 to 40 basis points. The bank is able to make this margin because of its advantages, compared with private banks, in having much lower default risk. IDA credits are interest-free but are subject to a commitment fee and a service charge which, in 2007, amounted to 95 basis points.

The Bank as Development Agency. The most basic rationale for development policy is to rectify market failures in the development process. In the early stages of development, these arise particularly because of the complementarity between activities, many of which become profitable only when other activities are already established. As a result of these failures, private sector enterprises may be unable to capture the social returns from potential investments. Governments can assist by undertaking investment projects in which private returns fall short of social returns. This can raise potential returns on private sector projects, thereby stimulating private sector investment.

Undertaking such policies would normally be identified as a role for domestic government. The bank, as a global, multilateral institution, has been involved in these activities partly because its knowledge and experience of development allow it to identify and address these needs more effectively than

can governments themselves. In many of the poorest of developing countries governments lack the capacity to operate such policy, and in addition, governments are often constrained by the politics of rent seeking. This government-failure view leads the bank to impose conditions on its loans and to monitor the projects in which it invests.

There are also global public-good aspects of good development policies. "Global commons" problems exist when a situation negatively affects many countries simultaneously. Examples are environmental pollution, the management of global capital flows, and the elimination of drug trafficking. Because the costs of global commons problems are so widespread, they require global solutions. The bank has played a significant role in work of this kind.

The Bank as Provider of Aid. Through its IDA arm, the bank carries out international development assistance directed toward the reduction of poverty. In doing this it provides concessional loans with a large aid component. Arguments for such aid rest on the idea of countries being stuck in a poverty trap from which escape is difficult without assistance from abroad (see Sachs 2005). The use of a multilateral institution such as the bank to provide these funds allows richer countries to give aid to poorer countries without this leading to direct bilateral political power relations with them. Channeling aid through a multilateral agency enables donors to precommit not to impose political conditions or economic conditions such as tying the aid to purchases of the donor country's goods.

The bank's ambition is to promote growth and alleviate the *causes* of poverty, rather than just to alleviate the *symptoms* of poverty. Reduction of poverty in this way came to the fore during the tenure of James Wolfensohn as bank president from 1995 to 2005. It is now widely accepted that poverty reduction through capability building is a necessary precondition for sustainable development.

Criticisms The bank has been heavily criticized by a wide range of people. Their criticisms generally fall into five categories. The first two criticisms are associated with left-wing critics and, more recently, with the anti-globalization movement. The third and

fourth have been associated with professional economists and the right wing of the political spectrum. The final criticism has been voiced by both political scientists and anti-globalization activists.

1. Is the Bank "Neoliberal"? The bank has been accused of imposing "neoliberal" economic policies on countries under the guise of structural adjustment. *Neoliberal* is a label widely attached by international relations scholars to those who see economic development as arising from the pursuance of free market policies (of the kind recommended by neoclassical economics) in conjunction with liberal, democratic political processes. John Williamson coined the term *Washington consensus* to describe those who think about economic development in this way. The majority of professional economists trained in British and U.S. universities over the past two decades would position themselves broadly within such a paradigm, and this is reflected in the bank's staff. Non-economists often see such a position as ideological; professional economists justify their position using both theory and evidence.

These debates are central to an understanding of the structural adjustment policies used by the bank since the 1980s. Prior to that time, the bank had lent almost exclusively for projects, but with the drying up of private flows to developing countries in response to the default by Mexico in 1982, and subsequently elsewhere in Latin America, countries' foreign exchange requirements outstripped the availability of attractive projects. On the "neoliberal" view that countries' abilities to respond were inhibited by over-large public sectors and lack of competition, the bank started to lend to finance reform (see Gilbert and Vines 2000; Mallaby 2004). The critics are correct in arguing that governments were pushed into adopting reforms that they would not otherwise have chosen. But most economists would argue that the critics are incorrect in suggesting that alternative policies might have enabled countries to cope with these problems with greater success or at lower cost. Evidence supporting this position comes, for example, from Tanzania, which attempted to go it alone until the mid-1990s but since reforming has seen much faster growth.

2. Conditionality The bank makes reform of policy a condition for its lending. Hopkins et al. (2000) set out the rationale for the bank's conditionality. The bank's status as a multilateral development agency enables it to impose these conditions, which are designed to improve outcomes. As described earlier, the bank's status allows it to lend at a rate that safeguards its debt servicing and repayment, reducing its risk premium and lowering costs of funds so that they are only marginally above its borrowing rate and less then the rate at which countries would be able to borrow from private markets. The imposition of conditions is a feature of the bank's operations that allows it to lend in circumstances that the private sector avoids.

Critics focus on how these conditions interfere with the sovereignty of the borrowing governments. If the critics' arguments were accepted and the bank was prevented from conditioning its lending on what it regards as appropriate policies, it would be unable to act as a financial intermediary in the way that it does.

But conditionality can lead to policy incoherence, since it results in time inconsistency. Recipient governments may agree ex ante to policy conditions if they see this as a means of attracting future loans or aid support. However, once the aid is disbursed these incentives are diminished, the reforms agreed to in the conditions will not be "owned," resulting in weak implementation, and in certain cases, governments may renege on the conditions to which they had previously agreed. The bank's response to a problem of this type has been to restructure its decisions as a repeated game by designing aid programs that include shorter horizons and more detailed conditions. But governments then have the incentive to hold back from reform and exaggerate its costs. An agile government may be able to sell the same reform several times (see Collier 2000).

As a result, there is a widespread view that bank policy conditionality has been not only undesirable but ineffective as well. Evidence comes in three forms   from case studies, from the bank's own assessments, and from econometric studies (see World Bank 2005 for a concise summary of the large liter-

ature). The failure of conditionality can be put down to a number of difficulties: implementation problems and the impact of external shocks (weather, commodity prices, etc.) that can throw plans off course, the overriding importance of domestic politics in determining policy and a failure of "ownership" of reforms, and the inadequacy of penalties and rewards that the bank can impose in the context of these political imperatives.

These negative findings have prompted the view that the bank should give up on the kind of policy conditionality described above (Burnside and Dollar 2000; Collier 2000). Instead the bank should recognize that there are "good" types and "bad" types of countries, and it should direct aid at the "good" types rather than trying to control the behavior of its clients. The bank should advocate and push for the adoption of policies that are consistent with its accumulated knowledge and practical experience. Governments will have an incentive to pay heed to this advice, because the bank bases its advice on its accumulated knowledge and because following this advice will induce future assistance from the bank: the bank should direct its lending and aid selectively toward countries that have demonstrated that they have good policies. A large component of bank activity in countries with poor policy environments would then become educational.

3. Aid Dependency The aid dependency thesis has been most cogently argued by Easterly (2006) and, in more qualified terms, by Collier (2007). It is an attack on the development agency role of the bank, or at least the way this role has been implemented hitherto. The bank lends to governments but, in Easterly's view, it is individuals and firms who promote development. Lending to governments reinforces the power of groups who have little to gain from development and who in any case can contribute little to that process. Collier particularly focuses on how bad governance and corruption can prevent development. He criticizes the provision of aid, both in countries where it is in danger of being used to further political conflict and in resource-rich countries in which the pursuit of resource rents can come to stifle development.

Both Collier and Easterly seek a substantial redirection of multilateral aid efforts. In part, this might mean a shift of resources within the bank toward the IFC, but it also requires new thought on what aid is effective. Much of the policy debate over the past two decades has been directed toward the question of "Where is aid effective?" We now need to ask, "In those countries in which aid has not hitherto been highly effective, what sort of aid may be effective?"

4. Lending to Emerging Market Economies Critics on the right have argued that the bank should withdraw from lending to middle income countries, because they are now able to access international capital markets on their own. These critics want the bank to retreat to being a development agency concerning itself purely with low-income countries that lack this access (see International Advisory Commission 2000). Ironically, moving in this direction would produce the same effects as those sought by left critics who wish to deny the bank the ability to impose policy conditionalities

The critics of bank lending to emerging markets fail to understand that without its lending to middle income countries, the bank would find its development assistance role in low income countries severely restricted the bank's current mode of operation depends on this lending. In the absence of a major increase in multilateral funding for development assistance, this financial intermediation function is required to generate the funds for development assistance. For example, in 2007, the bank received about $1 billion in income from its $40 billion of reserves. It also received profits on market-based lending of about $800 million and $800 million from fees charged to countries that borrow from the soft-loan IDA operations. If the bank were to cease market-based lending, then, even if it kept its reserves, it would forgo $800 million or so in annual revenue. According to Mallaby (2004, 409 10), that would not be enough to cover the administrative budget ($1.7 billion) plus the annual transfers that the bank now makes to the IDA ($300 million) and the HIPC debt relief fund ($240 million).

5. Representativeness The final set of criticisms relates to the bank's governance. The bank is governed by a 24-member Executive Board. Although board decisions are almost always consensual, the threat of a vote makes the distribution of votes important. This distribution in part reflects financial contributions to the bank's activities but even more, the status and power of countries in the 1944 Bretton Woods meeting. Uniquely, the United States has veto power on this board. European countries and Japan, together with the United States, make up a significant proportion of the voting strength. Sub-Saharan Africa, which is now the bank's most important client constituency, has the smallest share of votes 19 percent of IDA votes (2007) in relation to 48 percent of IDA credits (2002 7 average). By unwritten convention, the United States has the power to appoint the president, something which proved unpopular both when Paul Wolfowitz was appointed to replace James Wolfensohn in 2005 and when Robert Zoellick was appointed after Wolfowitz was forced to resign in 2007. There is now considerable pressure to reform the governance and voting structure of the organization, to make it more representative (see Woods 2006).

There is general agreement that both the board voting structure and the procedures for appointment of a president need to be overhauled, but as with the similar debates relating to the IMF and the UN Security Council, there is less agreement on the precise reforms, and in particular on which countries should be the beneficiaries of the reforms. The paramount danger in these reform proposals is that, in becoming more representative and democratic, the bank may also become less effective. Concretely, if a reformed bank were to advocate policies that donor country governments considered likely to be ineffective or counterproductive, they would reduce support for multilateral assistance. UNCTAD provides an example of a development institution in which all countries have equal voting power, and which has provided an effective forum in which developing countries can express their views. It has never enjoyed sufficient funding to have a major developmental impact.

The Bank as a "Bundled" Institution The bank has led global development research and has been

important in promoting debate, and fostering consensus, on the conditions and policies that will result in sustainable growth. A famous example is the role played by the bank in the early 1980s in the promotion of more open trade policies. This remains true in such diverse areas as poverty reduction strategies, governance, and financial regulation. Although few argue that the bank has originated major new ideas or directions in development economics, bank staff have generated a large volume of empirical research that has been fundamental in the assessment of the scale of problems (e.g., levels and changes in world poverty), the importance of different responses, and the applicability of possible policy reforms (e.g., on aging and HIV/AIDS) (see Squire 2000). World Bank (1998) brings together the results of a large body of important research in the bank on the effectiveness of aid. Gilbert and Vines (2000) refer to this as the "Knowledge Bank."

The bank's operation requires it to "bundle" together this knowledge of development work with its profitable lending to middle-income emerging market economies. This lending enables the bank to generate profits to support its other activities. We can think of the bank as rather like a large private American university, which uses its endowment income and income from lending to subsidize its research and teaching (see Gilbert and Vines 2000).

Squire (2000) clearly describes the value of this bundling: "Without an in-house capacity, integrating the results of research into the World Bank's everyday operations and making those results available to policy-makers in developing countries does not happen. This usually requires an in-house champion, and the best champion is usually the best researcher." Thus the bank brings both research and money. Accountability in relation to the expenditure of public money forces research applicability. And the bank's research culture ensures that projects are directed more closely to developmental and poverty reduction objectives than would otherwise be the case: money carries ideas.

It is hard to see how else this bundling of knowledge and money could be provided. In private banks, or private consulting companies, there will always be cost-minimizing pressures to cut the knowledge overhead, subject merely to satisfactory product (i.e., research) quality. Furthermore private banks have no incentive to disseminate knowledge and make their experience widely available, and consulting firms do so only at a price substantially in excess of the marginal cost of dissemination. By contrast, university research departments produce knowledge, but not necessarily of the kind needed for good policy. In addition knowledge about best-practice development policy is a global public good, requiring a global solution to the provision-problem of the kind that the bank provides. Many reforms of policy that work in one country will tend to work elsewhere; there are significant economies of scale and scope in the development of knowledge about development. It is difficult for individual governments to provide these goods.

Gilbert and Vines (2000) argue that such bundling can also help one to better understand bank conditionality. Conditionality may be necessary if the bank is to be able to lend where commercial banks are unwilling to enter. But the bank's knowledge is necessary to determine what constitutes good development policies and also to identify conditions for the reform of policies of a kind that governments will be willing to "own."

The conception of the bank as a Knowledge Bank is central to this account of what the bank does. The bank's lending provides income from endowment and revenue earned in the provision of loans. This funds the activities of the bank as a development agency, including the production of research that increases the effectiveness of both the lending and the aid that the bank provides.

As emphasized earlier, it is important that countries get the *right* combination of money and knowledge. In the countries with good policies and strong institutions the emphasis should be on money. In countries where the policy environment is poor, either because of an unsatisfactory macroeconomic and trade framework or because of poorly performing public institutions, the emphasis must be on getting these policy and institutional prerequisites right.

See also aid, international; evolution of development thinking; structural adjustment; Washington consensus

FURTHER READING

Burnside, C., and D. Dollar. 2000. "Aid, Growth, the In centive Regime, and Poverty Reduction." In *The World Bank: Policies and Structure*, edited by C. L. Gilbert and D. Vines. Cambridge: Cambridge University Press, chapter 8. Argues that aid should be directed to countries with good policies.

Collier P. 2000. "Conditionality, Dependence, and Policy Coordination: Three Current Debates in Aid Policy." In *The World Bank: Policies and Structure*, edited by C. L. Gilbert and D. Vines. Cambridge: Cambridge University Press, chapter 12. Critical of the effectiveness of conditionality as practiced by the World Bank.

———. 2007. *The Bottom Billion*. Oxford: Oxford University Press. Examines traps due to conflict, the natural resources "curse," geography, and poor governance, and argues that aid ought to be used to enable countries to escape from these traps.

Easterly, W. 2006. *The White Man's Burden*. Oxford: Oxford University Press. Argues against a view that aid should be used in a big development push and in favor of a smaller amount of carefully targeted aid toward those who might use it best.

Gavin, M., and D. Rodrik. 1995. "The World Bank in Historical Perspective." *American Economic Review, Papers and Proceedings* 85 (2): 329 34. Useful historical account up until the mid 1990s.

Gilbert C. L., and D. Vines, eds. 2000. *The World Bank: Policies and Structure.* Cambridge: Cambridge University Press. An edited volume discussing how the bank's policies could be redirected toward it becoming a "Knowledge Bank" and how the bank's structure could be reformed to assist with this.

Hopkins, R., A. Powell, A. Roy, and C. L. Gilbert. 2000. "The World Bank, Conditionality, and the Compre hensive Development Framework." In *The World Bank: Policies and Structure*, edited by C. L. Gilbert and D. Vines. Cambridge: Cambridge University Press, chapter 11. Sets out the argument for conditionality and exam ines difficulties with it.

House, B., D. Vines, and W. M. Corden. 2008. "The In ternational Monetary Fund." *New Palgrave Dictionary of Economics*. A discussion of the history of the IMF and its involvement in the resolution of financial crises, inter national macroeconomics surveillance, and the devel opment of macroeconomic policies in the world's poorest countries.

International Advisory Commission. 2000. *Report of the International Financial Institution Advisory Commission.* Washington, DC: Government Printing Office. The report of a commission, chaired by Allan Meltzer, critical of the IMF and the World Bank and calling for their activities to be scaled back.

Kapur, D., J. Lewis, and R. Webb. 1998. *The World Bank: Its First Half Century.* Washington, DC: Brookings In stitution. A careful and thorough history of the bank.

Mallaby, S. 2004. *The World's Banker: A Story of Failed States, Financial Crises, and the Wealth and Poverty of Nations.* New Haven, CT: Yale University Press. A bi ography of James Wolfensohn, president of the bank from 1995 to 2005, which is at the same time a gripping account of what the bank does and its history.

Sachs, J. 2005. *The End of Poverty.* New York: Penguin. Makes the case for a very large increase in aid in order to help lift the poorest countries out of poverty.

Squire, L. 2000. "Why the World Bank Should Be Involved in Development Research." In *The World Bank: Policies and Structure*, edited by C. L. Gilbert and D. Vines. Cambridge: Cambridge University Press, chapter 4. Argues that involvement in development research is central to the bank's mission of promoting good devel opment policy.

Wolfensohn, J. 1999. "A Proposal for a Comprehen sive Development Framework A Discussion Draft." Washington, DC: World Bank. Available at http:// www.worldbank.org/cdf/cdf text.htm. Written by the former head of the bank, this document describes a set of principles to guide development and poverty reduction.

Woods, N. 2006. *The Globalizers*. Ithaca, NY: Cornell University Press. A revealing study of the relation ships between the IMF, the World Bank, and their clients, and of the political processes that lead policies to be adopted.

World Bank. 1998. *Assessing Aid: What Works, What Doesn't, and Why?* Oxford: Oxford University Press. A systematic account of the possibilities and difficulties in the provision of aid.

———. 2005. *Review of World Bank Conditionality: The Theory and Practice of Conditionality, a Literature Review.* Washington, DC: World Bank.

———. 2006. *Annual Report.* Washington, DC: World Bank.

———, Operations Evaluation Department. 2005. *Improving the World Bank's Development Effectiveness: What Does Evaluation Show?* Washington, DC: World Bank.

CHRISTOPHER L. GILBERT AND DAVID VINES

■ World Economic Forum

The World Economic Forum (WEF) is one of the major instances of informal power influencing the world economy. Incorporated as a Swiss not-for-profit foundation whose members are among the world's 1,000 largest companies, the WEF is best known for the annual meetings, attracting around 3,000 participants, that it organizes in Davos, Switzerland, where chief executive officers (CEOs) meet world leaders, government officials, and a broad range of media, cultural, and academic executives identified as opinion leaders. It also organizes yearly meetings with more regional focuses, as well as ad hoc meetings related to the various initiatives it hosts under its Center for Public-Private Partnership.

The WEF has its origins in the Centre d'Etudes Industrielles of Geneva, one of the four executive business schools that had a key role in the managerial revolution following the Marshall Plan in postwar Europe. A junior faculty member, Klaus Schwab, was asked to prepare the 25th anniversary celebration of the institution, which was planned for 1971. He suggested celebrating by organizing a large conference in Davos mingling U.S. experts in management with European businessmen. He also conceived an artful legal artifice to keep control of the event and future initiatives. Thanks to the worldwide reputation of the Centre d'Etudes Industrielles, the conference attracted big names from American business schools and high-profile European industrialists. It is against this background that Klaus Schwab is known as the founder and president of the WEF. The scope of the forum quickly expanded beyond managerial questions to develop an agenda including macroeconomic, political, and strategic issues.

Since the early 1990s, the large increase in the number and prestige, as well as in the geographical and sectoral origins, of the participants has transformed the WEF into the world rallying point for which it is now known. The number of participants grew from fewer than 1,000 in 1971 to around 3,000 by 2007, more than half of whom are invited by the foundation to participate in panels, lend a high profile to the network, and ensure the greatest possible impact. While regular members pay around $20,000 a year to take part in the events, several dozen WEF partners pay more than ten times this amount to be entitled to direct input into the agendas of the meetings by setting up panels, choosing participants, and providing strategic positioning for their industries. Guests include several hundred political leaders and high-ranking officials, including around 20 chiefs of state or government each year; ministers; governors of central banks; directors of international organizations; Forum Fellows chosen from among academic scholars; executive officers of research foundations; some social, religious, and cultural celebrities; as well as around 500 editors and columnists from large press groups around the world. The apogee of the prestige of participants to the forum was probably reached in 2000 when for the first time the president of the United States, William J. Clinton, took part in the event with nine other political leaders of the Organisation for Economic Co-operation and Development (OECD).

Although the WEF advertises itself as an international organization committed to improving the state of the world by engaging leaders in partnerships on key issues to shape global, regional, and industry agendas, critics are prone to identify the WEF as nothing but a meeting place for world masters planning neoliberal globalization behind closed doors. Of course, those closely associated with the forum are inclined to deny its power, while those fiercely opposed are likely to emphasize

its overarching influence. More broadly, however, these contrasting views of the influence and power of the WEF in global politics and economy express disagreement on one outstanding feature of the changes associated with globalization: the significance of the power exercised on a global scale by informal and weakly institutionalized nonstate actors.

The WEF presumably reflects the archetype of the most exclusive and powerful transnational elite club. Other groups include the Bildeberg conferences, the Trilateral Commission, the Clinton Global Initiative, or on a more regional basis, the U.S.-based Business Council and the European Round Table of Industrialists. The WEF is an extremely efficient infrastructure for piling up in three days appointments that would otherwise take three months, solving urgent matters among numerous actors, adding up pressure on the public agenda, and ensuring a large impact in elite circles, the media, and wider public opinion. Situated at the interface between economic and political aspects of the world order, it also embodies a strategic function in seeking to find a synthesis between the territorial basis of a legal and political order and the global reach of the world economy. This is why, for instance, key actors in the Middle East conflict such as Mahmoud Abbas, president of the Palestinian Authority, and Simon Peres, top Israeli politician for decades, can mingle with Eric Schmidt, cofounder and CEO of Google, or Matthew J. Cadbury, heir of the British Cadbury Schweppes agribusiness company. The WEF meetings and backup services provided by its secretariat play a crucial role not only in developing common strategies but also in raising collective consciousness among elites from different countries who get better acquainted and influence one another.

The contribution of elite groups like the WEF to the structure of power in the world economy takes two interrelated courses. On the one hand, they reinforce the formation of an overall class consciousness that transcends national and industry-based rivalries and embraces the global dimension of capitalism. On the other hand, they constitute a privileged locus for expanding this worldview to subordinated interests. Media and other dissemination channels target issues easily identified with the general interest. It is against this background that pleas for historic changes from top women and men in Davos should be understood. The WEF, however, faces intrinsic limits to its power. Its actual practices and organizational principles impair its power. It will always lack a recognized institutional basis. The bigger its suggested influence, the greater the problems it faces in legitimizing its informal power. The sheer size of the meetings, the huge media coverage, and the at times better-organized opposition requiring massive security measures have also tarnished the public image of the club atmosphere that made the WEF so successful.

Despite its continued attractiveness to world leaders for reinforcing their class consciousness, shaping the global agenda, and cutting deals, the WEF has tried to overcome these internal and external challenges since the early 2000s. Officials of the foundation have given a twist to its activities in order to bridge the gap between the loose influence of a closed club and a more focused engagement with states, intergovernmental organizations, and society at large in supporting the launch of various initiatives. Foundation officials and guests target a large range of commitments, from health to the digital divide, including education, gender, corruption, religion, energy challenges, and global warming. Although many of the participants drift apart eventually, some of them vie with other initiatives to concretely influence the policy agenda.

One of the self-proclaimed flagships of the WEF in this regard is the GAVI Alliance (Global Alliance for Vaccines and Immunization). Created as a global network of international development organizations, national governments, multilateral development banks, philanthropic organizations, and large corporations involved in vaccines and immunization, it was officially launched at the 2000 annual meeting at the initiative of the former World Health Organization director-general Gro Harlem Brundtland, in association with Bill Gates, CEO of Microsoft and

cofounder of the Bill and Melinda Gates Foundation, and other regular participants to the WEF meetings. In December 2006, the amount actually disbursed since its creation reached $1.3 billion.

The Global Competitiveness Report is another well-advertised product of the WEF. It was created in 1979 in conjunction with the International Institute for Management Development, the top European executive business school located in Lausanne, Switzerland. The report uses a large quantity of data to build an index that ranks states according to the environment they provide for supporting the economy. The yearly release of the report in advance of the WEF annual meetings makes headlines. Although not at the same level as the conditionality imposed by international financial institutions on developing countries or the recommendations included in the OECD country and sectoral reports for industrialized countries, WEF competitiveness reports still operate at a distance. By shifting the focus away from management issues related to firm performance to competitiveness concerns in national business environments, they explicitly fulfill an agenda-setting function and contribute to harnessing state capacity to define overall economic orientations.

In conclusion, the WEF has achieved long-standing prominence among the various organizations designed to provide informal platforms for meetings, networking, lobbying, and strategic planning among an impressive list of world leaders. Like other platforms, the WEF claims wide public recognition on various issues as a forum of transnational private governance. Yet its loose and large influence is not without limits, as its so-called club atmosphere has faded away with the growth of its events, and it will always offer a privileged target for protesters and all those contesting its legitimacy. The future of the WEF clearly lies in the narrow path between persisting with its massive communication campaign supporting its public image and returning to a more intimate format that would preserve the discrete atmosphere for business leaders to meet with one another and with government officials.

See also anti-globalization; globalization; Group of Seven/ Eight (G7/G8)

FURTHER READING

Berne Declaration. www.publiceye.ch. Web site of a coalition of nongovernmental organizations publicly engaging the WEF.

Graz, Jean Christophe. 2003. "How Powerful Are Transnational Elite Clubs? The Social Myth of the World Economic Forum." *New Political Economy* 8 (3): 321 40. A scholarly account from an international political economy perspective, which relies on archival material and around 50 interviews.

Lapham, Lewis. 1998. *The Agony of Mammon*. London: Verso. A scathing account of two WEF annual meetings by the editor of *Harper's* magazine.

Pigman, Geoffrey. 2006. *The World Economic Forum.* London: Routledge. The only exhaustive, though uncritical, published monograph on the WEF.

Public Citizen's Global Trade Watch. 2002. *Davos World Economic Forum: Pricey Corporate Trade Association Loses Its Camouflage*. Washington D.C.: Public Citizen. A well researched report from the division of the U.S. national consumer and environmental group promoting government and corporate accountability in the globalization and trade arena.

World Economic Forum. www.weforum.org. The official Web site of the WEF.

JEAN CHRISTOPHE GRAZ

■ World Intellectual Property Organization

The World Intellectual Property Organization (WIPO) is the key international organization dealing exclusively with intellectual property rights. As such, and with more than 180 member states, it plays a pivotal role in the negotiation and establishment of intellectual property norms. It was founded in 1970 under the terms of a convention signed in Stockholm in 1967. Under a 1974 agreement, WIPO became a specialized agency of the United Nations.

History The origins of the organization lie in the 1893 merger of the secretariats (or "international bureaus") of the Paris and Berne unions, which made

up the contracting parties to the two oldest and most important intellectual property treaties: the 1883 Paris Convention for the Protection of Industrial Property and the 1886 Berne Convention for the Protection of Literary and Artistic Works. The merged organization was known as the Bureaux Internationaux Réunis de la Protection de la Propriété Intellectuelle (BIRPI).

The idea of transforming BIRPI into an international intellectual property organization arose at a 1962 meeting of the Permanent Bureau of the Paris Union and the Berne Union. The meeting recommended the setting up of a Committee of Governmental Experts in order to consider administrative and structural reforms to the Paris and Berne Union systems and prepare for a diplomatic conference.

It is important to note that, during this time, the decolonization process that began just after World War II was gathering pace. Many new, developing countries were becoming independent and seeking to join the United Nations and other international organizations. The United Nations itself was undergoing a period of transformation as it sought to accommodate a rapidly increasing membership with a wide range of interests and concerns. Which parts of the UN system should have jurisdiction over complex and politically contentious matters such as intellectual property rule-making remained unresolved. In consequence, it was obvious that BIRPI could no longer remain as a developed-country "club" and needed a more multilateral character that could attract developing countries, including the newly independent ones.

Some developing nations had their own ideas about international intellectual property norm-setting, however, and were becoming assertive in expressing them. This was cause for concern in some quarters. Indeed, the proposal to establish a new organization based on BIRPI was intended in part to ensure that politicized organizations, including those known for accommodating the specific concerns of the developing countries, would not be chosen as the forum for negotiating intellectual property norms.

According to Stephen P. Ladas (1975), the intent was "to head off any attempt by outsiders, such as the United Nations Economic and Social Council or the United Nations Conference on Trade and Development, to deal with the subject of intellectual property and eventually to form a Specialized Agency of the United Nations in this field."

A second meeting of the committee took place in 1966 and was attended by representatives from 39 nations, of which 9 were developing countries. The rest were developed or European communist countries. The draft convention prepared by BIRPI on the basis of the views expressed by the committee at these two meetings was presented to a diplomatic conference in 1967 in Stockholm, where a final text was approved. The WIPO Secretariat, located in Geneva, Switzerland, is still known as the International Bureau.

Activities, Governance, and Mandate As of late 2007, WIPO administered 24 multilateral agreements. These were of three kinds:

1. *The standard-setting treaties*, which define agreed basic standards of protection for the different intellectual property rights. These include the Paris Convention; the Berne Convention; the 1961 Rome Convention for the Protection of Performers, Producers of Phonograms, and Broadcasting Organizations; the 1996 WIPO Copyright Treaty; and the 2006 Singapore Treaty on the Law of Trademarks.

2. *The global protection system treaties*, which facilitate filing or registering of rights in more than one country. These include the 1970 Patent Cooperation Treaty, the 1891 Madrid Agreement Concerning the International Registration of Marks, and the 1958 Lisbon Agreement for the Protection of Appellations of Origin and their International Registration.

3. *The classification treaties*, which are essentially administrative and concern the management and organization of information concerning patents, trademarks, and

industrial designs. These include the 1957 Nice Agreement Concerning the International Classification of Goods and Services for the Purposes of the Registration of Marks and the 1971 Strasbourg Agreement Concerning the International Patent Classification.

WIPO has three main governing bodies, which are the General Assembly, the Conference, and the Coordination Committee. There are also standing committees and working groups. It is an intergovernmental organization, but official observers are allowed to attend meetings subject to their compliance with certain criteria and procedures. These observers include other intergovernmental organizations and nongovernmental organizations. The latter category comprises not only groups representing business, intellectual property owners, and legal practitioners, which tend to favor intellectual property protection and support WIPO, but also groups whose views range from the very pro intellectual property to the extremely critical.

What is WIPO's mandate? The organization's two objectives as stated in Article 3 of the convention are: (1) to promote the protection of intellectual property throughout the world through cooperation among states and, where appropriate, in collaboration with any other international organization; and (2) to ensure administrative cooperation among the unions. However, the 1974 agreement between the United Nations and the World Intellectual Property Organization balanced the goal of promoting intellectual property protection by recognizing WIPO's responsibility "for facilitating the transfer of technology related to industrial property to the developing countries in order to accelerate economic, social and cultural development, subject to the competence and responsibilities of the United Nations and its organs."

Recent Challenges WIPO briefly lost influence between 1986, when intellectual property was included in the Uruguay Round of trade negotiations, and 1994, when the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) became part of the Agreement Establishing the World Trade Organization, and also for a short time afterward. But since then, WIPO has become a high-profile actor, raising levels of intellectual property protection in developing countries and driving the evolution of international intellectual property lawmaking in response to the opportunities and challenges arising from new technologies. For example, WIPO has proffered technical assistance to developing countries implementing TRIPS and provided a forum for the negotiation of new intellectual property treaties, including four since 1996. In addition, it has become a forum for substantive discussion of emerging intellectual property issues, including ones of interest to developing countries, such as protection of traditional knowledge and cultural expressions. Significantly, it is now a well-funded organization on account of the fees charged for services it provides to the private sector.

From time to time, WIPO is criticized for being biased, either in favor of the interests of the developing countries or in favor of those of industry and the developed countries. In the past few years, complaints have been made that it promotes intellectual property protection in ways that benefit transnational corporations but do not suit developing countries. For example, some critics, including governments, claim that WIPO's technical assistance programs for developing countries give insufficient attention to public interest exceptions to intellectual property rights.

Since 2004, 14 developing country WIPO members, led by Brazil and Argentina, have sought to establish a Development Agenda at WIPO to address the perceived bias. These countries, collectively known as the Friends of Development, argue that as a UN agency WIPO should be guided by the Millennium Development Goals, adopted by the UN General Assembly in 2000. To the 14 countries, the agenda should cover such issues as WIPO's mandate and governance, norm-setting, technical cooperation, and transfer of technology. As of late 2007, the agenda had not yet been approved. Nonetheless, the organization has had to position itself publicly as one whose work is dedicated to fostering economic development.

WIPO plays a substantial role in the world economy by dint of its involvement in promoting intellectual property rights internationally. The high commitment of so many governments and businesses to influence negotiations at WIPO is testament to the high economic stakes involved in the regulation of intellectual property rights at the global level. Because of the more obvious economic significance of the World Trade Organization, however, and the huge attention given to the TRIPS Agreement, WIPO tends to receive less attention from trade analysts than it probably should.

See also Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS); intellectual property rights

FURTHER READING

Ladas, Stephen P. 1975. *Patents, Trademarks, and Related Rights: National and International Protection.* Vol. 1. Cambridge, MA: Harvard University Press. Contains a useful and informed historical view of the origins and antecedents of WIPO and clarifies the political context for the creation of the organization from the view of a person who was directly involved.

May, Christopher. 2007. *The World Intellectual Property Organization: Resurgence and the Development Agenda.* London: Routledge. The only book written specifically on WIPO. It explains how WIPO operates, describes the role of WIPO within contemporary global politics, and discusses its relationship with the World Trade Organization and the TRIPS Agreement.

Menescal, Andréa K. 2005. "Changing WIPO's Ways? The 2004 Development Agenda in Historical Perspective." *Journal of World Intellectual Property* 8: 761 96. Well researched account of WIPO's recent engagement in development related intellectual property issues from a historical perspective.

Musungu, Sisule F., and Graham Dutfield. 2003. "Multi lateral Agreements and a TRIPS Plus World: The World Intellectual Property Organization." TRIPS Is sues Paper No. 3. Geneva: Quaker United Nations Office. Available at http://www.quno.org/geneva/pdf/economic/Issues/Multilateral Agreements in TRIPS plus English.pdf. Controversial critical examination of the institutional aspects of WIPO, including the internal architecture of the organization and the allegedly ex cessive influence of legal practitioner organizations and business associations.

UN Conference on Trade and Development and Interna tional Center for Trade and Sustainable Development. 2003. *Intellectual Property Rights: Implications for Development.* Geneva: UNCTAD ICTSD Capacity Building Project on IPRs. Available at http://www.iprsonline.org/unctadictsd/policyDpaper.htm#policy. Written for an audience with a general interest in the subject, it describes the international intellectual prop erty regulatory system and explains the central position of WIPO within it.

GRAHAM DUTFIELD

■ World Trade Organization

The World Trade Organization (WTO) is *the* international institution governing virtually all trade among its 151 members. As such, it serves two primary functions. First, it provides a set of multilaterally agreed rules that govern policies affecting (1) countries' trade in goods and services and (2) the protection of intellectual property of foreign nationals in member countries. By imposing certain disciplines on countries' trade-related policies, the rules covering trade in goods and services encourage international trade by making future access to foreign markets more predictable. Second, it provides a forum for (1) the day-to-day administration of the various WTO agreements, (2) the settlement of trade disputes among members, and (3) negotiations to further liberalize trade as well as to strengthen and extend the multilateral rules.

Between 1947 and 2007, first under the General Agreement on Tariffs and Trade (GATT), and then under the WTO, world merchandise trade expanded markedly faster than world production (6.2 percent a year on average versus 3.8 percent). It is no coincidence that this period of increasing international specialization, facilitated by the two rules-based multilateral agreements, was also a period of unparalleled prosperity for the world economy, with average per capita income growing more than twice as

fast (2.1 percent a year versus 0.9 percent) as in the previous "boom" period of 1820 1913.

Around the time the WTO came into being, the growing share of output traded across national frontiers, coupled with expanded flows of foreign investment and technology, gave rise to what has become the economic catchword of our times: *globalization* (Wolf 2004). With the world's economies progressively more integrated the essence of globalization and economic prosperity more and more dependent on developments in world trade, the WTO is both the preeminent international economic organization and, inevitably, the international organization most closely linked in the public's mind with globalization.

Origin of the WTO After more than seven years of often arduous negotiations, the Uruguay Round the eighth and final round of trade negotiations under the auspices of the GATT ended successfully in December 1993. In April 1994 trade ministers meeting in Marrakesh, Morocco, signed the Uruguay Round Agreements, including the agreement establishing the WTO, and on January 1, 1995, the WTO came into being. With the subsequent entry of China an event of major importance for the global economy Chinese Taipei (Taiwan), and Saudi Arabia, and with the expected entry of Russia, the WTO's members will include all of the major trading nations (at the end of 2007 an additional 28 countries were in various stages of the accession process).

Although the WTO grew out of the GATT and embraces nearly all of the GATT's principles and practices, it is a distinctly different entity. In contrast to GATT's status as a provisional agreement among governments, the WTO is a formal international organization. The rules, which under the GATT covered only trade in goods, were extended to cover trade in services (telecommunications, banking, tourism, etc.) and the protection of intellectual property. Whereas the GATT had included limited membership agreements dating from the Tokyo Round negotiations (1973 79), all the major WTO agreements are binding on all the members.

The GATT, updated by the Uruguay Round, was absorbed by the WTO and became the umbrella treaty for trade in goods (the new agreement is referred to as GATT 1994, the original one as GATT 1947). As noted earlier, the principles that guided the original GATT, along with virtually all of the key rules and procedures of the original GATT, can be found in the WTO.

Basic Principles Five time-tested basic principles carried over from the GATT some explicit, some implicit inform the rules and guide the activities of the WTO.

Nondiscrimination. The most-favored-nation (MFN) clause prohibits a WTO member from discriminating between imports of goods or services from other WTO members. By requiring a country to treat imports from all WTO members equally, the MFN rule depoliticizes trade (it also ensures that a country buys its imports from the cheapest foreign source). In the case of services (but not goods), countries could obtain at the inception of the services agreement, or on accession for new members, "MFN exemptions" to cover measures that would otherwise breach the MFN rule.

Some key exceptions to the MFN principle included in the GATT were carried over to the WTO, most notably the provisions for (1) free trade areas and customs unions among small groups of countries and (2) preferential treatment of imports from developing countries. With the rapid spread of preferential trading arrangements especially free trade areas beginning in the late 1980s, many trade experts openly question the practical relevance of the MFN rules (see, for example, Srinivasan 2005).

While the MFN clause enforces nondiscrimination at the border, the national treatment principle prohibits discrimination between imported *goods* and domestic products once the imports have entered the country. A retail sales tax applied to imported automobiles but not to domestically produced automobiles, for example, would violate the national treatment principle (however, imposing an import duty on foreign-made automobiles as they enter the country is not a violation of national treatment). For imported *services*, in contrast, application of national treatment is required only to the extent that a country has undertaken an explicit commitment to do so on the service in question.

Tariffs only. The WTO is not a free trade organization. Member countries are not required to participate in trade liberalizing negotiations, and they are free to protect domestic producers by imposing restrictions on imported goods and services. In the case of goods, however, if they do protect they are required to use tariffs only rather than import quotas or other nontariff restrictions to provide the protection. This is good economics tariffs are a less distorting (less inefficient) way of granting protection, and the level of protection provided by a tariff is more transparent than that provided by nontariff barriers.

Consensus decision making. Although there are provisions for majority voting on a one-country, one-vote basis, the WTO has continued the strong tradition, developed under the GATT, of making decisions by consensus defined as no country present at the meeting when the decision was taken formally objecting to the proposed decision. Although this often slows the decision-making process, it has important advantages, including adding to the intellectual and diplomatic "legitimacy" of the decisions, many of which result in legally binding obligations for all members.

Special treatment for low-income countries. Following a practice that evolved under the GATT, the WTO rules and procedures include special arrangements for developing countries. These include special provisions in WTO agreements (for example, longer implementation periods for new obligations), activities of the Committee on Trade and Development, and technical assistance (mainly training). Responding to a widening income gap among these countries a number of successful "developing" countries now have per capita incomes approaching those of developed countries, while many others remain very poor the WTO uses a two-part grouping: developing countries and least-developed countries (a UN grouping of the 50 poorest countries, 32 of which are WTO members), coupled with an increasing focus on helping the latter group of countries (see Part Five of WTO 1999).

Small Secretariat with a limited mandate. The member countries have also continued the tradition, developed under the GATT, of having a Secretariat that is *very* small in terms of budget and staff as compared with the other major international economic organizations, such as the World Bank and the International Monetary Fund (IMF). The limited mandate is reflected not only in the limited budget and staff, but also, for example, in the fact that the Secretariat is not allowed to interpret WTO rules, nor can the Director-General initiate a dispute settlement case, no matter how blatant the rule violation. More generally, the head of the WTO has much less authority in relation to both the member countries and the organization's activities than do his counterparts at the World Bank and the IMF.

The result is a Secretariat whose activities are largely limited to servicing meetings and other activities of the member countries. This "member-driven" nature of the WTO, which has become even more pronounced than it was under the GATT, can be traced to the fact that trade policies are highly politicized at the national level. No government especially no government of a large or medium-size country can be seen advocating a Secretariat with any real independence and power. (Some analysts argue that the member-driven nature of the GATT was an important factor in its success; see Hudec 1993; Winham 1998.)

Administering the Agreements and Other Day-to-Day Activities Aside from the agreement establishing the WTO, virtually all of the more than 50 agreements, annexes, decisions, and understandings agreed to in the Uruguay Round, as well as the organization's day-to-day activities, fall into one of five broad areas: administering the agreements on trade in goods, on trade in services, and on intellectual property protection, plus dispute settlement and periodic reviews of countries' policies under the Trade Policy Review Mechanism (TPRM). Technical assistance and training activities for officials from low-income countries round out the list of usual activities.

The member-driven nature of the work, plus the fact that all of the nearly 30 councils, committees, working parties, and working groups are open to all WTO members (there are no limited membership "executive groupings"), means that national

delegations in Geneva typically have to cope with a heavy schedule of meetings dealing with a variety of often technically complex issues. This can be burdensome for delegations from the smaller developing countries and especially so for the least-developed countries (Blackhurst, Lyakurwa, and Oyejide 2000).

Dispute Settlement A rules-based agreement can function only if it includes an effective system for resolving disputes between the member countries. Although the GATT's dispute settlement system worked well for much of nearly five decades that the GATT provided the legal framework for world trade, by the early 1980s serious problems were beginning to appear. In particular, there was a noticeable decline in the willingness of countries to "protect" the system by (1) allowing the dispute settlement process to run its course (under GATT rules, any member including the defendant country could block the process at any stage) and (2) removing policies that were found to be inconsistent with the country's GATT obligations. Everyone agreed that negotiations to reform the dispute settlement system were a high-priority part of the Uruguay Round agenda.

While most of the GATT's rules and procedures were taken over by the WTO with relatively little or no substantive modification, an important exception to this generalization occurred in the Uruguay Round understanding on dispute settlement. The previous practice of requiring that decisions by the panels that hear cases be adopted by consensus was turned on its head under the WTO, panel decisions are automatically adopted unless there is a consensus *not* to do so. A second key change was the creation of a new institutional element the Appellate Body to which the parties to a dispute can appeal the panel's decision, with the Appellate Body's decision being final. Under the heading "Strengthening of the Multilateral System," the understanding also explicitly prohibits the use of *unilateral* actions in trade disputes between WTO members.

The WTO's dispute settlement system was remarkably successful during its first decade, producing more than 150 panel and Appellate Body reports which, in turn, generated a rich case law. While not disputing that assessment, many countries especially developing countries believe the system can be improved and have offered specific proposals for reform (WTO 2004, chapter VI; Sacerdoti, Yanovich, and Bohanes 2006).

Negotiating Rounds The underlying "ethos" of the WTO is that the pursuit of *freer* trade is a desirable goal. During the GATT years that goal was pursued through a series of formal negotiating rounds (eight in total). Despite some support for the view that more progress might be possible through lower profile "continuous" negotiations, during its first decade the WTO embraced the big round approach with the launch in Doha, Qatar, in November 2001 of the Doha Development Agenda, encompassing both a new round of trade liberalizing negotiations and work on other issues, especially problems many developing countries were encountering in implementing existing commitments agreed to in the Uruguay Round.

At the end of 2007 the outlook for a successful conclusion of the Doha Round remained clouded. This, in turn, has given rise to claims that a failure of the round would seriously, perhaps fatally, harm the WTO. As a ploy to get countries to make the politically difficult decisions needed to conclude the round successfully, there is nothing objectionable about such claims. At the same time, they overlook at least two important sources of strength and stability in the WTO system. First, existing WTO rules and contractual obligations, together with the well-functioning dispute settlement system, impose severe constraints on the ability of countries to *increase* current levels of protection. Of course countries could be tempted to flout the WTO, but then the second source of strength comes into play, namely the fact that politically powerful multinational corporations and other export-oriented firms from many countries, together with their millions of employees, have a very large financial stake in world trade continuing to flow smoothly and be governed by enforceable multilateral rules.

Other Challenges Facing the WTO An international organization with a mandate as broad and complex as that of the WTO is bound to find itself

confronting a seemingly endless series of challenges. Very briefly, these include:

- *Dispute settlement.* It was noted earlier that, although the dispute settlement system is working very well, a number of proposals for improving the system are on the table. In addition, there is a fear that a failure of the Doha Round could put additional strains—perhaps very serious strains—on the dispute settlement system if countries sought to obtain through litigation what could not be achieved through negotiation—for example, the reduction or elimination of agricultural subsidies. It is helpful to recall, in this context, that the WTO has no police force and no jails. The dispute settlement system works only if the member countries want it to and are willing to protect it.

- *Preferential trading arrangements.* The continuing proliferation of preferential trading arrangements, in particular free trade areas, is believed by many to be a serious problem for the WTO. However, 20 years of debate and analysis have failed to bring about a consensus among either economists or policymakers on whether, in practice, such agreements—many of which extend to issues not covered by the WTO—support or harm the multilateral trading system (see WTO 2004, chapter II; WTO 1995). It is widely assumed that a failure of the Doha Round would give an added stimulus to the spread of such agreements.

- *Decision making.* The WTO's various councils, committees, and negotiating groups are loosely analogous to what in a national government would be the "parliamentary" branch of government. It is generally accepted in the trade policy community that the practices and procedures in this area—especially the ones surrounding the negotiating activities—are functioning badly and need to be reformed. Along with built-in procedural and structural shortcomings, other factors that have been blamed for the disappointing pace of negotiations and decision making include the increased WTO membership, the increased sensitivity of the issues being considered, the lack of a more regular presence in Geneva of senior policymakers from capitals, and the across-the-board reliance on consensus decision making (see WTO 2004, chapters VII and VIII).

- *Anti-globalization.* The WTO is a major target of the anti-globalization movement—the high point or low point, depending on one's perspective, being the "Battle of Seattle" at the December 1999 Ministerial meeting when an estimated 40,000 to 50,000 people showed up to protest a very wide range of issues (Jones 2004). Aware that such attacks on the WTO can erode public support for further trade liberalization, the WTO has taken a number of steps in the years since 1999 to counter what it believes to be the misconceptions and distortions spread by the anti-globalization movement. These include improving the transparency of WTO activities and the dialogue with nongovernmental organizations (NGOs), expediting the derestriction of documents, providing regular briefings for NGOs in Geneva, holding symposia, and allowing NGO attendance at ministerial meetings (see WTO 2004, especially chapter V; also chapters I and III).

The WTO's Contribution Douglass North shared the 1993 Nobel Prize in economics for his work on the role of institutions in early European and American economic growth and in the experience with postwar growth in developing countries. His focus was on human cooperation that allows societies to capture the gains from specialization and trade stressed by Adam Smith in the *Wealth of Nations* (North 1990). He argues convincingly that an important determinant of a country's rate of economic growth is the extent to which institutions, as they evolve, create or fail to create a favorable environment for cooperative solutions to complex economic exchange.

Although North's analysis of differences in growth rates among countries, and within the same country over time, focused on the role of institutions at the subnational and national levels, there is little doubt that global economic integration spurred by ever more complex production and distribution chains has progressed to the point that his findings apply with equal force at the international level. If economic growth is a priority, a global economy requires multilateral economic institutions that can perform the essential institutional functions that, in successful economies, are carried out at the national level by national institutions. The WTO is the preeminent example of such a multilateral institution.

See also Agreement on Agriculture; Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS); agricultural trade negotiations; anti-globalization; competition policy; customs unions; Doha Round; free trade area; General Agreement on Tariffs and Trade (GATT); General Agreement on Trade in Services (GATS); global public goods; globalization; government procurement; labor standards; multilateralism; nondiscrimination; nontariff measures; quotas; regionalism; sanitary and phytosanitary measures; special and differential treatment; tariffs; technical barriers to trade; trade facilitation; trade in services; Trade Policy Review Mechanism; trade-related capacity building; Uruguay Round; World Trade Organization, accession to; World Trade Organization dispute settlement

FURTHER READING

Barton, John, Judith Goldstein, Timothy Josling, and Richard Steinberg. 2006. *The Evolution of the Trade Regime: Politics, Law, and Economics of the GATT and the WTO.* Princeton, NJ: Princeton University Press. An excellent up to date interdisciplinary overview of the WTO regime.

Blackhurst, Richard, Bill Lyakurwa, and Ademola Oyejide. 2000. "Options for Improving Africa's Participation in the WTO." *The World Economy* 23 (4): 491 510. Reprinted in *Developing Countries and the WTO: A Pro active Agenda*, edited by Bernard Hoekman and Will Martin. United Kingdom: Blackwell, 2001. Examines the challenges of WTO participation confronting virtually every WTO member in sub Sahara Africa (as well

as low income countries elsewhere), including sugges tions for reallocating existing budget resources to im prove the situation.

Goldstein, Judith, Douglas Rivers, and Michael Tomz. 2006. "Institutions in International Relations: Under standing the Effects of GATT and the WTO on World Trade." *International Organization* 61: 2. Assesses the trade implications of membership in the GATT and the WTO and provides empirical evidence that membership increased trade for both developed and developing countries.

Goode, Walter. 2007. *Dictionary of Trade Policy Terms.* 5th ed. Cambridge: Cambridge University Press. A nearly indispensable guide to terms used in the jargon filled world of the WTO and other international economic organizations. Also covers various WTO rules, newer issues, and developing country concerns

Hudec, Robert E. 1993. *Enforcing International Trade Law: The Evolution of the Modern GATT Legal System.* Salem, NH: Butterworth Legal Publishers. A standard reference work, by a leading GATT legal scholar, on the GATT legal system as it stood just prior to its major overhaul in 1995 as part of the Uruguay Round Agreements.

Irwin, Douglas A. 2005. *Free Trade under Fire.* 2d ed. Princeton, NJ: Princeton University Press. An excellent review of the case for free (freer) trade that also examines the various arguments for protection. Chapter 2 contains a nontechnical summary of the various ways trade lib eralization stimulates economic growth.

Jones, Kent. 2004. *Who's Afraid of the WTO?* Oxford: Ox ford University Press. After cataloging a long list of in terest groups that have been highly critical of the WTO, the author carefully analyzes and refutes virtually all of the charges leveled against the WTO. The author also provides an excellent overview of the WTO's contribu tion to economic and noneconomic goals worldwide.

Krueger, Anne O., ed. 1998. *The WTO as an International Organization.* Chicago: University of Chicago Press. A conference volume containing 15 papers by leading WTO scholars from economics, law, and political sci ence covering both institutional and substantive issues and challenges.

North, Douglass. 1990. *Institutions, Institutional Change, and Economic Performance.* Cambridge: Cambridge University Press. Perhaps the best summary of the work

that earned him the Nobel Prize in Economics in 1993, this book presents an analytical framework for analyzing how institutions, by influencing the ability of humans to cooperate in the economic sphere, play a key role in economic growth.

Petersmann, Ernst Ulrich, ed. 2005. *Reforming the World Trading System: Legitimacy, Efficiency, and Democratic Governance.* Oxford: Oxford University Press. A collection of 29 essays by leading WTO specialists from the fields of economics, law, and trade diplomacy covering virtually every issue that has been raised in policy debates over reforming the WTO system as it enters its second decade.

Sacerdoti, Giorgio, Alan Yanovich, and Jan Bohanes, eds. 2006. *The WTO at Ten: The Contribution of the Dispute Settlement System.* Geneva and Cambridge: WTO Secretariat and Cambridge University Press. A collection of articles by a diverse group of experts including academics, policymakers, judges from a number of international tribunals, and members of the WTO's Appellate Body written in connection with events commemorating the Appellate Body's 10th anniversary.

Srinivasan, T. N. 2005. "Nondiscrimination in GATT/WTO: Was There Anything to Begin with and Is There Anything Left?" *World Trade Review* 4 (1): 69 95. The author argues that the exceptions to nondiscrimination that were in the GATT from the beginning have become dominant over time to the point that little nondiscrimination remains today. He also considers the relevance of nondiscrimination to the new areas covered by the WTO services and the protection of intellectual property.

Winham, Gilbert R. 1998. "The World Trade Organization: Institution Building in the Multilateral Trade System." *The World Economy* 21 (3): 349 68. A political scientist's view of the operation of the GATT and its implications for the WTO.

Wolf, Martin. 2004. *Why Globalization Works.* New Haven, CT: Yale University Press. A well written and detailed analysis of the case for globalization that covers all aspects of the topic. The reference of choice for an in depth introduction to the forces shaping the world economy.

World Trade Organization. 1995. *Regionalism and the World Trading System.* Geneva: WTO Secretariat. Although the statistics on the number of preferential agreements in existence are not up to date (it's a fast moving area), the rest of the study, including the historical part and the analytical parts which offer a detailed analysis of the many often subtle ways that preferential agreements and the WTO system can interact, for good and for bad remain fully relevant.

. 1999. *Guide to the Uruguay Round Agreements.* Geneva: WTO Secretariat. In terms of rhetoric and organization the official legal texts of the Uruguay Round results are virtually incomprehensible to the nonexpert. This book by WTO Secretariat staff members presents the results in an authoritative but understandable way that also makes it easy to get a good sense of the overall package.

. 2004. *The Future of the WTO: Addressing Institutional Challenges in the New Millennium.* Geneva: WTO Secretariat. The four core chapters of this report by an independent group of eight international experts provide an informed overview of current institutional challenges in the areas of dispute settlement (VI), decision making (VII and VIII), and the role of the Director General and the Secretariat (IX). A free copy in English may be downloaded from http://www.wto.org/english/thewto e/10anniv e/future wto e.pdf. Also available in French and Spanish.

. 2005. *Understanding the WTO.* 3d ed. Geneva: WTO Secretariat. A useful, reasonably detailed introduction to both the WTO system and current challenges confronting the system. At 112 pages it occupies a space between brief descriptions (such as this encyclopedia entry) and detailed academic work on each of the many rules and activities of the GATT/WTO. An up to date free version in English may be downloaded from http://www.wto.org/english/thewto e/whatis e/tif e/understanding e.pdf. Also available in French and Spanish.

. http://www.wto.org. A comprehensive user friendly Web site relevant to all aspects of the WTO's history and current activities.

RICHARD BLACKHURST

■ World Trade Organization, accession to

Policymakers give a range of reasons for wanting their countries to join or accede to the World Trade

Organization (WTO). For some the rationale is to further integrate their country into the world economy. Here often the hope is that the better access to foreign markets that WTO membership can bring will result in higher exports. Another economic rationale is to encourage greater foreign direct investment and, more generally, to use WTO accession as a seal of approval recognized by the international business community. It is also the case, however, that many nations join the WTO for political reasons. Transition economies, for example, often see WTO membership as a means to join the international community of nations. Some see WTO accession as facilitating both political as well as economic reform processes within their countries.

There is some overlap between these stated rationales and the potential benefits of WTO accession identified in economic research, in particular as they relate to bolstering exports and foreign direct investment inflows. Economists would also point to the benefits that flow from better foreign access to the acceding nation's markets, specifically in terms of lower prices for and a greater variety of imports. By binding national tariffs, committing to eliminate quotas on imports, and reforming other state measures, an acceding nation can enhance the credibility of its policies and reduce the uncertainty facing the private sector. In principle, then, WTO accession can improve important components of the national business environment which, in turn, has sizable domestic payoffs.

The Process of WTO Accession Most WTO members were also contracting parties to the General Agreement on Tariffs and Trade (GATT) and automatically became founder-members of the WTO when it was established on January 1, 1995, because they had either signed the Marrakesh Agreement Establishing the World Trade Organization in April 1994 or because they had joined the GATT after April 1994 but before the WTO was set up. The Marrakesh Agreement also made provision for "any state or customs territory possessing full autonomy in the conduct of its external commercial relations and to the other matters provided for in this Agreement" to accede to the WTO. From January 1, 1995, to July 2007, 23 new members, including two least-developed countries (LDCs), have acceded to the WTO. At the time of writing, another 28 jurisdictions, 10 of which are LDCs, are in the process of acceding to the WTO.

Article XII.1 of the Marrakesh Agreement permits new members to accede on "terms to be agreed" between the applicant and the WTO. No guidance is given in the Marrakesh Agreement on the "terms to be agreed." However, the WTO Secretariat, in consultation with WTO members, has drawn up a set of procedures for accession that are closely modeled on those followed by contracting parties to the GATT. In brief, these procedures require that:

1. an applicant must send a communication to the Director-General of the WTO indicating its desire to accede to the WTO under Article XII;
2. this communication is then circulated to all WTO members and a decision is made whether to approve the establishment of a working party;
3. once a working party has been established and a chairperson has been appointed, the applicant submits a memorandum on its foreign trade regime;
4. the working party meets to examine the memorandum and the applicant provides further information and answers;
5. a Working Party Report is prepared;
6. at the same time, bilateral negotiations are undertaken on concessions and commitments on market access for goods and services (as well as other specific terms of accession);
7. a goods schedule and a services schedule are prepared;
8. thereafter, a draft decision and a draft Protocol of Accession (containing commitments listed in the Working Party Report and the goods and services schedules) are prepared;
9. the working party, and then the WTO's General Council or the Ministerial Conference, approves the accession package;

10. thereafter, the applicant formally accepts the accession package;
11. the applicant notifies the WTO Secretariat of its formal acceptance; and
12. 30 days later the applicant becomes a member of the WTO.

For the countries that joined the WTO most recently this process took on average 10 years.

The "Price" of WTO Accession Throughout the WTO accession process the onus is on the applicant to satisfy the demands of existing WTO members. This apparently one-sided procedure has given rise to the following perceptions about the accession process (Evenett and Primo Braga 2005):

- The WTO accession process is costly and complex.
- The WTO accession process is taking longer and longer to complete.
- The price of joining the WTO now includes commitments that go beyond the GATT/WTO agreements.
- The price of joining the WTO is steadily rising.
- The WTO accession process takes little account of the specific circumstances of applicant countries or their needs for special and differential treatment.

With respect to the evidence on the "price" of accession, it is important to distinguish between the two broad types of commitments made by acceding countries: those relating directly to market access (for goods and services) and other commitments on rules. Acceding countries may benefit from transition periods or exceptions to existing WTO rules, but these are rarely granted by existing WTO members. Concerning commitments to open national markets to foreign trade, in the areas of agricultural and non-agricultural (typically manufacturing) products there is clear evidence that the price of accession is growing over time. For both agricultural and nonagricultural goods the average tariff binding that acceding countries were allowed has fallen over time (to approximately 15 and 6 percent, respectively) and is now at levels well below those agreed by developing countries in the Uruguay Round (Evenett and Primo

Braga 2005). In short, from a mercantilist perspective the relative price of WTO accession is high (in comparison with Uruguay Round commitments made by peer nations) and growing over time.

From the first 20 completed WTO accessions the picture that emerges concerning service sector commitments tells a similar story. Taking the number of services subsectors (of the 160 identified in the WTO's classification list) committed by countries as a proxy for the "price" to be a WTO member, LDCs that were founding members of the WTO committed on average 20 subsectors. The averages for founding members in the developing and industrialized country categories, in turn, were respectively 44 and 108. Countries that have acceded since 1995, in turn, have on average committed around 104 subsectors. Needless to say, this is a crude measure of the "services-related" price of accession as this figure does not capture the depth (e.g., the scheduling of explicit limitations) or the breadth (e.g., the modes of delivery covered) of the commitment. Still, it is illustrative that countries that went through the WTO accession process typically committed a much higher number of subsectors than GATT contracting parties at a similar level of development did in the context of the Uruguay Round negotiations.

Turning to commitments on other policies that countries have adopted when joining the WTO, the picture is more mixed. With the exceptions of China and Taiwan accession countries signed around 25 such commitments. These commitments typically concern a wide range of state measures, some of which are not obviously trade related. Bulgaria, for example, made commitments with respect to domestic price controls, the privatization of state-owned enterprises, and excise taxes on alcohol, as well as many other traditional trade policy related measures. A controversial question is whether this class of accession obligation goes beyond the commitments agreed during the Uruguay Round (constituting so-called WTO-plus commitments) or involves an accession country agreeing to forgo the rights available to other WTO members (the so-called WTO-minus commitments).

Whether an accession obligation goes beyond an existing WTO agreement depends in large part on how the latter is interpreted, and so it should not be surprising that disagreement is rife on the extent of WTO-plus commitments. Moreover, some WTO-plus obligations may only involve consultation with, or reporting to, existing WTO members and thus are of limited developmental significance. Others may be more significant, such as Jordan's commitment that if any of its laws or state acts are subsequently found to contradict international treaties (not just WTO agreements), then the latter would have precedence. WTO-minus commitments are easier to identify, such as Ecuador's commitment to eliminate all subsidies before the date of accession and its commitment never to introduce them afterward. China's acceptance of product-specific transitional safeguard provisions, which can be more easily triggered than regular WTO safeguards, provides another example. WTO-plus and WTO-minus commitments differentiate WTO members, and they could be interpreted as contributing to a multitier multilateral trade system. This systemic concern is in addition to any of the adverse developmental effects that may result from these specific commitments.

In sum, there is evidence that the accession process is becoming more demanding in terms of market access commitments. In July 2004 WTO members recognized this trend and agreed that less liberalization be demanded of newly acceded members in the agriculture and nonagricultural market access negotiations during the Doha Round. Whether there is a trend increase in WTO-plus or WTO-minus commitments is unclear, but the very fact that existing WTO rules allow for them is a source of concern. Evidence on the latter should be interpreted with care in view of the possibility of a potential bias in the sample of recently acceded countries. After all, most of these countries were transition economies with highly distorted trading regimes to start with. Accordingly it could be argued that the higher demands of WTO members simply reflect this reality rather than a systemic trend.

The Empirical Studies of the Impact of WTO Accession The critical welfare question, however, is not whether the price of WTO accession is rising, but whether the price is worth paying in terms of its developmental impact. If it is, then the demands made by existing WTO members of acceding countries might be characterized as "tough love." Otherwise, the WTO accession process may well be seen as a one-sided power play whereby current WTO members wring commercial advantage out of weaker economic partners.

A comprehensive evaluation of WTO accession should examine postaccession performance on many metrics and should consider the state measures taken before and after the date of WTO accession. At present, few accession countries have five or more years of postaccession data to begin identifying the effect of WTO accession, so the available evidence here is necessarily limited.

The available country-specific research on WTO accession has almost exclusively focused on the case of China, an exception being an analysis of the response of Bulgaria and Ecuador's exports to their accessions (Evenett, Gage, and Kennett 2004). Other country analyses typically refer to countries in the process of accession, and a recent paper by Lissovolik and Lissovolik (2004) on Russia's potential WTO accession is a good example of this type of analysis. These authors found that after controlling for customary determinants of bilateral trade flows Russia's exports to WTO members from 1995 to 2002 were smaller than those to nonmembers. Rather than immediately attribute the export underperformance to Russia's nonmembership in the WTO, Lissovolik and Lissovolik consider four other potential explanations for this finding. Only one such explanation (the nature and extent of Russian export controls) survives scrutiny and the authors qualify their conclusions accordingly.

The second class of relevant literature includes a series of academic studies that have called into question widely held and long-standing views about the economic effects of WTO membership. Having documented numerous claims made by scholars,

officials, and international organizations about the beneficial consequences of WTO membership (and of membership in its predecessor, the GATT), Andrew Rose used conventional econometric tools to assess these claims in a series of papers. In one study he found no statistically significant effect of GATT/WTO membership on the value of bilateral trade flows over a 50-year period, a finding he regards as mysterious (Rose 2004). In another study he found that GATT/WTO accession increased the value of bilateral trade flows, but WTO membership itself does not (Rose 2005a). Moreover GATT/WTO membership was found in a third study not to have a statistically significant effect on the volatility of trade flows, as measured by the coefficient of variation of bilateral trade flows over a 25-year period (Rose 2005b). In these three studies GATT/WTO membership is captured in econometric specifications by the use of dummy variables, rather than directly identifying the changes in trade policies that followed from membership or accession.

Rose's findings have prompted others to revisit the questions he posed and the methods he employed. Subramanian and Wei (2003) diverge from Rose's earlier articles in two respects. First, they differentiate between the effects of GATT/WTO membership on industrialized and developing countries. More generally they note that there is no presumption that the effects of GATT/WTO membership are common across all goods and trading partners, hence calling into question the pooling of certain types of trade data. Second, they employ some of the latest techniques for estimating bilateral gravity equations that try to control for country-specific characteristics. These two departures appear to have led to substantially different findings. Subramanian and Wei find that WTO membership raised trade between industrial countries by 40 percent, whereas such membership only increased a developing country's trade if it joined the WTO after the completion of the Uruguay Round. Asymmetric results were found across sectors, and GATT/WTO membership is associated with greater effects on goods trade (which typically faces lower trade barriers than other

sectors). The authors note that these findings are consistent with the reluctance of industrialized economies to liberalize trade in food and clothing under the auspices of the GATT and WTO, and the fact that developing countries were not asked to undertake many liberalizing commitments before the completion of the Uruguay Round.

Noneconomists have also contributed to the literature on the effects of WTO accession. Goldstein, Rivers, and Tomz (2003) draw on the insights of the international relations literature to assess the effects of international institutions on trade patterns. They note that many customs territories (which need not be nations) participated in the GATT and WTO before formally becoming members of these institutions. The date of accession, therefore, may not accurately reflect the point at which a nation begins to align itself with multilateral trade rules. Another interesting observation is that the effects of GATT and WTO membership are likely to be conditional on prevailing preferential trading arrangements and on the nature and history of certain state-to-state ties, such as former colonial links. These authors also undertake a gravity equation analysis and find that, although the GATT/WTO appears to have increased trade among members, these institutions have contributed to faster growth of trade among nonmembers.

To summarize, there is a dearth of empirical evidence on the effects of WTO accession on the developmental prospects of developing countries. This weak evidential base has created a vacuum that has been filled by doubts about the case for opening markets and taking WTO obligations and negotiations seriously. Some have questioned whether developing economies, especially least-developed countries, should be encouraged to use the WTO as a forum to structure national development efforts. Moreover it is difficult to see how technical assistance programs, especially those initiatives that enable firms to capitalize on postaccession export opportunities, can be effectively designed without a better understanding of the economic consequences of WTO accession. One challenge for the scholarly

and policymaking communities in the coming years is to fill this knowledge gap.

See also General Agreement on Tariffs and Trade (GATT); political economy of trade policy; trade and economic development, international; World Trade Organization

FURTHER READING

Evenett, Simon J., Jonathan Gage, and Maxine Kennett. 2004. "The Impact of Joining the WTO on Market Access: Detailed Evidence from Bulgaria and Ecuador." Mimeo (September). Oxford University, Saïd Business School. An empirical evaluation of the different poten tial effects of WTO accession on two nations.

Evenett, Simon J., and Carlos A. Primo Braga. 2005. "WTO Accession: Lessons from Experience." *Trade Note 22*. International Trade Department. Washington, DC: World Bank. A summary of the principal findings on WTO accession taking a deliberately developmental perspective.

Goldstein, Judith, Douglas Rivers, and Michael Tomz. 2003. "How Does the Trade Regime Affect Interna tional Trade?" Mimeo. Stanford University, Depart ment of Political Science. An empirical analysis of the effects of different international institutions on inter national trade flows.

Kennett, Maxine. 2007. *The Principal Legal Obligations of Accession to the World Trade Organization.* Manuscript. Geneva: Ideas Center. A comprehensive analysis of the legal obligations taken on by acceding nations in the first 12 years of the WTO's existence.

Lissovolik, Bodgan, and Yaroslav Lissovolik. 2004. "Russia and the WTO: The 'Gravity' of Outside Status." Mimeo. Washington, DC: International Monetary Fund. An analysis of the potential effect of WTO ac cession on Russian trading patterns.

Rose, Andrew. 2004. "Do We Really Know that the WTO Increases Trade?" *American Economic Review* 94 (1): 98 114. A much cited paper that finds no statistically sig nificant effect of the WTO or GATT membership on national trade flows.

———. 2005a. "Which International Institutions Promote International Trade?" *Review of International Economics* 13 (4): 682 98. An empirical analysis of the effects of membership of different international institutions on trade flows.

———. 2005b. "Does the WTO Make Trade More Stable?" *Open Economies Review* 16 (1): 7 22. An em pirical examination of the effects of WTO membership on the stability of trade flows.

Subramanian, Arvind, and Shang Jin Wei. 2003. "The WTO Promotes Trade, Strongly but Unevenly." Working Paper No. WP/03/185. Washington, DC: International Monetary Fund. A paper that counters Rose (2005), arguing that a disaggregated empirical analysis generates very different findings.

World Trade Organization. 2005. *Technical Note on the Accession Process.* WT/ACC/10/Rev.3 (28 November). Geneva: WTO. A long, detailed statement of various matters relating to the accession of countries to the WTO.

———. http://www.wto.org/english/thewto e/acc e/acc e .htm. A Web site with the WTO's resources relating to accession.

SIMON J. EVENETT AND CARLOS A. PRIMO BRAGA

■ World Trade Organization dispute settlement

The dispute settlement system of the World Trade Organization (WTO) is the mechanism through which members thought to be in violation of their trade agreement commitments present their cases to a third-party panel for decision. The system evolved from the dispute settlement process of the WTO's predecessor, the General Agreement on Tariffs and Trade (GATT), which first embarked on the process of resolving disputes among its signatories in 1948. GATT's dispute settlement process was initially re ferred to as a process of "conciliation," not arbitra tion, and was supervised by a panel of experts that generally consisted of diplomats, not lawyers.

The goal of this process was to reach a solution mutually agreeable to the parties rather than to es tablish and apply binding legal principles. It did this reasonably well, but there was one major shortcom ing. GATT, as its name makes clear, was a multi lateral agreement. Indeed, strictly speaking, GATT had no "members." Signatories were known as "Contracting Parties." Since GATT was simply a

contract among a group of parties, any action required a consensus of all of those parties. This allowed parties whose practices were challenged to block the establishment of a dispute settlement panel or the adoption of an adverse panel report. There was considerable dissatisfaction with this aspect of the system.

Although some amendments to improve the GATT dispute settlement system were made over the years, it took the establishment of the WTO's dispute settlement system in 1995 to put in place an effective legal system, including an Appellate Body to which panel decisions may be appealed.

The WTO dispute settlement system is governed by the Understanding on Rules and Procedures Governing the Settlement of Disputes, known as the "Dispute Settlement Understanding" or "DSU," and is administered by a Dispute Settlement Body (DSB), which consists of representatives of each WTO member. The DSB is the sole WTO body with authority to establish panels, to adopt their reports and those of the Appellate Body, to maintain surveillance of the implementation of the rulings and recommendations it adopts, and to authorize the suspension of concessions or other obligations under WTO agreements if its rulings and recommendations are not acted on by members in a timely fashion.

The power of the DSB is, perhaps, the biggest difference between the GATT and WTO systems. Consensus is still required, but it is "reverse consensus." Consensus is required *not* to establish a panel when a request is made, reversing the GATT practice of requiring consensus to do so. Since at least one member, the complaining party, always wants the panel to be established, this effectively means there can be no reverse consensus. This reverse consensus rule also applies to the other major functions of the DSB, including the adoption of reports and the authorization of the prevailing member to impose the WTO remedy of suspension of trade concessions or other obligations on a losing member that does not implement the DSB's recommendations.

Jurisdiction The WTO agreements are between governments and impose obligations only on gov-

ernments. Therefore access to the dispute settlement system is limited to member governments   private interests have no right of access. Although panels and the Appellate Body sometimes accept amicus curiae submissions from private interests, they have rarely accorded much weight to these submissions in their analysis.

The subject matter jurisdiction of panels and the Appellate Body is limited to the interpretation of the provisions of the WTO agreements cited by the complaining party. Although there is much discussion of whether WTO panels and the Appellate Body should consider other international agreements and principles of public international law as part of their work, they have done so only to assist in interpreting the WTO agreements. The DSU instructs panels to interpret the agreements in accordance with the "customary rules of interpretation of public international law." This has been understood to refer to the Articles 31 and 32 of the Vienna Convention on the Law of Treaties, which places primary emphasis on the role of the ordinary meaning of a treaty's text rather than the history of its negotiation.

The Panel Process The WTO panel process is, unexpectedly, a legalistic one, with rules providing for the establishment and composition of panels and for the procedures they are to follow. The system's diplomatic heritage in GATT is reflected at the outset, however, by the requirement that the process begin with a formal request from one member to another for consultations concerning the subject of the dispute. Only after these are held may a member with a grievance move on to request the establishment of a panel and its procedures to resolve the dispute.

Consultations The consultation requirement concerns those "measures affecting the operation of any covered agreement taken within the territory" of one member that are deemed by another member to contravene WTO commitments. In most cases, however, informal bilateral consultations between the parties have already taken place; consequently, the formal request for consultations is often seen as an effort by the requesting member to signal to the other member that the matter at issue is serious.

The purpose of consultations is to enable the parties to gather relevant, and correct, information both to assist them in reaching a mutually agreed solution or, failing that, to assist them in presenting accurate information to the panel. Many disputes do not proceed beyond the consultations stage.

The request for consultations plays an important role in defining the jurisdiction of the eventual panel in the dispute. The DSU requires that the complaining party give the reasons for the request, identify the measures at issue, and indicate the legal basis for the complaint. This forms the basis for a subsequent request for the establishment of a panel, although there need not be an exact identity between the measures identified in the request for consultations and those identified in a request for the establishment of a panel. Consultations are confidential and take place bilaterally without any WTO involvement. What transpires during the consultations is not the concern of a panel.

Other WTO members have a limited role in consultations. A complaining member may draft its request for consultations so that even third-party members with a substantial trade interest may not participate. Tactical considerations usually dictate the complaining member's decision on this point. Of course, those third-party members may participate in the panel procedure itself even if they have not been permitted to participate in consultations.

Requests to Establish a Panel The complaining member's request for the establishment of a panel is a fundamental, jurisdictional document. It is incorporated into the panel's terms of reference and, therefore, determines the subject matter jurisdiction of the panel. It also fulfills the due process objective of notifying the responding party   and other WTO members as potential third parties   of the claims at issue. Because of its jurisdictional importance, disagreements over the scope of terms of reference arise frequently.

The formal requirements of a request for the establishment of a panel are set out in the DSU: the request must be in writing, indicate whether consultations were held, identify the objectionable measure taken by the other member, and specify the

"legal basis for the complaint." A measure typically is a law or regulation, or an action that applies a law or regulation. The cited measure, together with the legal basis for the complaint, that is, the provisions of the relevant WTO agreement cited in the panel request, constitutes the "matter" over which the panel has jurisdiction.

Composition of Panels Panels normally are composed of three persons, with one serving as chair. Panelists usually are present or former members of nonparty delegations to the WTO or academics who are not paid for their service. Unless the parties agree, panelists may not be nationals of parties or third parties to the dispute. They serve in their individual capacities, not as government representatives or representatives of other organizations. A developing country that is a party to a dispute may request that at least one panelist shall be from a developing country.

Usually panelists are nominated by the WTO Secretariat from member delegations and from a roster of governmental and nongovernmental individuals who are deemed qualified by virtue of their training or experience. If the parties do not agree on panelists within 20 days of the establishment of the panel, either party may request the Director-General of the WTO to name the panelists. Although the DSU instructs members not to oppose nominees of the Secretariat except for compelling reasons, parties frequently have proven unable to agree on the composition of the panel, leaving the Director-General to appoint the panel.

Procedures Panels normally issue their own detailed working procedures soon after their members have been named. These procedures supplement the more general Working Procedures set out in the DSU and may cover topics ranging from the panel's schedule and deadlines to the required format for submissions. Although the DSU sets out some time limits for panels' work, these often prove too short, and panels frequently depart from them, usually with the agreement of the parties and notice to the DSB. Despite these delays, WTO dispute settlement is widely seen as comparatively rapid for litigation as complex as most WTO disputes are.

Although normally all of the relevant information is submitted to a panel by the parties, panels are not confined to the parties as sources of the information they need. They have the right to seek information and technical advice from any source they deem appropriate. They may obtain opinions from experts and, with respect to scientific and technical matters, may request an advisory report from a more formal expert review group.

Other WTO members may participate to a limited extent in panel proceedings as third parties. When a panel is established, the chairman of the DSB usually specifies that members wishing to participate as third parties should file their notifications within 10 days. Third parties are normally entitled to receive only the first submissions of the parties to the panel and to participate in a special session of the first meeting of the panel with the parties. However, members have been granted enhanced third-party rights in several instances.

Meetings of WTO panels with the parties are normally confidential, but, with the agreement of the parties, some panels have opened their meetings to the public or permitted the rebroadcast of portions of the hearings that do not involve confidential information.

Few individuals have served on more than one or two WTO dispute settlement panels. Accordingly, the role of the WTO Secretariat, which provides legal and secretarial advice to panels, cannot be overstated. Not all panelists feel qualified to disagree with the Secretariat on points of law. Although there have been proposals to create a permanent roster of panelists, these proposals have not yet achieved consensus approval.

Panels normally conduct two separate meetings with the parties. At these meetings, the parties present formal statements of their arguments or rebuttals, after which the panel questions them on the details of their arguments. Panels vary greatly in their degree of activity during meetings with the parties. Some will interrupt presentations to ask questions as they arise; others will question after a presentation is concluded; still others do so only after both sides have finished their presentations; and a few present no oral questions at all, but rely entirely on written questions.

Prior to the issuance of their final report, parties are given the opportunity to review an interim report. They may submit comments on this report and may request a meeting with the panel to discuss these comments. In most instances, parties do not request a meeting with the panel, but simply submit written comments on the report. The interim review stage is not intended to present an opportunity to reargue the issues. Generally panels do not revise their substantive findings following the interim review stage, although there have been some exceptions. Although interim reports are supposed to be strictly confidential, their release to the parties provides the first disclosure beyond the Secretariat of the outcome of the proceeding. The gist of the panel's findings frequently makes its way to the press at this point.

The Appellate Process The creation of a standing Appellate Body, staffed by its own secretariat and therefore largely independent of the rest of the WTO, was one of the two most important additions to the GATT dispute settlement system by the WTO. The other, of course, was "reverse consensus." About half of the reports of panels are appealed to the Appellate Body.

The WTO Appellate Body consists of seven members appointed for four-year terms, which may be renewed once. Members serve on a paid, part-time basis. The DSU requires that members be persons of recognized authority and demonstrated expertise in law, international trade, and the subject matter of the WTO agreements. Members may not be affiliated with any government. For the most part, the members have been experienced trade negotiators and lawyers. Generally they have been appointed with a view to ensuring a broad geographic representation on the Appellate Body.

Cases are heard by a Division of three of the seven members of the Appellate Body who are selected on the basis of a rotation system intended to ensure random selection and unpredictability. Unlike panelists, Appellate Body members may serve on any case, including those involving their own country. The Division selects a presiding member, whose

responsibilities include coordinating the overall conduct of the proceeding, chairing oral hearings and meetings, and coordinating the drafting of the appellate report. The Appellate Body has adopted its own working procedures, which specify the schedule for each proceeding as well as the required format and content of notices of appeals and other submissions. The Appellate Body's proceedings, like those of panels, are confidential and closed to the public.

Reports of the Appellate Body are signed by all members of the Division. The DSU provides that opinions expressed by individuals shall be anonymous. At the time of this writing (in late 2007), there have been two separate (dissenting) opinions contained in reports of the Appellate Body.

The jurisdiction of the Appellate Body is limited to issues of law covered in a panel report and to legal interpretations developed by a panel. The Appellate Body may uphold, modify, or reverse the legal findings and conclusions of a panel. If it modifies or reverses the findings, however, the Appellate Body may not remand the matter to the panel for further proceedings.

Perhaps the greatest impediment to the work of the Appellate Body is the requirement that it must normally complete the entire appeal process within 60 days and, according to the DSU, in not more than 90 days. Since the panel process normally averages approximately a year to complete, this is an extremely restrictive deadline. On occasion the Appellate Body has found it simply impossible to complete the task within this deadline and the parties have agreed to extend this deadline.

Adoption and Implementation of Reports Panel reports are formally adopted by the DSB within 60 days of their circulation to all members unless a party gives notice of its decision to appeal the report or the DSB decides by consensus not to adopt the report (which has never happened since the creation of the WTO). Appellate Body reports are normally adopted within 30 days of circulation, again unless the DSB decides by consensus — which must include the prevailing party — not to do so.

Generally, where they find in favor of a complaining party, panels and the Appellate Body simply recommend that the defending party bring its measure into conformity with the relevant WTO Agreement. The defending party must normally state its intentions in this regard at a DSB meeting within 30 days of the adoption of the report. The defending member then has a "reasonable period of time" within which to bring its measure into conformity. Ideally the duration of this "reasonable period of time" is negotiated between the parties to the dispute. If agreement cannot be reached, the DSU provides for an arbitration process, to be completed within 90 days, as to the duration of this period. These arbitrations are conducted by a sole arbitrator agreed on by the parties or, absent agreement, appointed by the WTO Director-General. Arbitrations are relatively infrequent.

In implementing the panel or Appellate Body rulings, the defending member need only ensure that its measures and actions are WTO-consistent with *prospective* effect. There is no obligation to take any action to redress the past effects of any WTO-inconsistent actions.

If the defending member fails to act within the reasonable period of time or the complaining member is not satisfied with the manner in which the defending member has acted, the complaining member may refer the matter back to the original panel. Following a truncated version of the normal process, the reconstituted panel issues a further report — nominally within a 90-day deadline but in practice within about 6 months — as to whether the defending member properly implemented the original panel or Appellate Body's report and is now in compliance with the relevant provisions of the WTO Agreements.

The DSU contains a procedural inconsistency in this area that has led to the so-called sequencing problem. In addition to providing for review of disputes concerning a defending member's compliance with rulings in adopted panel or Appellate Body reports, the DSU also grants complaining members rights (described later) to take certain retaliatory actions in the event of inadequate implementation. The deadline for both actions begins when the time for the defending member to comply has expired,

however. The DSU allows, as noted, 90 days for a review of compliance. But it calls for the retaliatory action to be taken within 30 days of the expiration of the time for the defending member to comply 60 days before the members can even know what the compliance ruling will be. In other words, a member wishing to retaliate is called on to do so 60 days before it can know whether it is entitled to do so. This "sequencing" problem has generally been resolved by bilateral agreements between the parties suspending and preserving the retaliatory rights of the complaining member pending the outcome of the review proceedings.

Remedies Where a defending member fails to implement properly an adverse panel ruling, both the GATT and WTO dispute settlement systems provided for remedies in the form of compensation or suspension of concessions. *Compensation*, in GATT/ WTO parlance, refers not to monetary compensation, but to a reduction in the defending member's tariff rates or other trade barriers designed to offset the trade harm to the complaining member resulting from the impugned measure. *Suspension of concessions* refers to the imposition by the complaining member of additional tariffs or other trade barriers against imports from the defending member.

The DSU provides that the level of suspension of concessions shall be "equivalent" to the level of nullification or impairment of benefits due to the complaining. This can be difficult to quantify, so the DSU provides the possibility of arbitration, by the original panel or an arbitrator appointed by the WTO Director-General, to determine the appropriate level of retaliation. Economists have frequently pointed out that these "remedies" do not benefit the member taking retaliatory action. Others have pointed out that they also do not benefit the complaining member's exporting industry, which presumably was the intended beneficiary of the action. Both of these points are correct, but they overlook the political statement that retaliatory action, or the credible threat of such action, makes in both the domestic politics of the members concerned and in the internal politics of the WTO. It is also noteworthy that, on several occasions, members have

obtained the right to retaliate but have refrained from doing so, presumably because they have realized that not only would such action not benefit their economies but, on the contrary, could harm them.

WTO members continue to negotiate to revise and improve the system. The issues under negotiation largely concern procedural matters, such as the sequencing problem and the panel selection process. Some members see the Appellate Body's lack of authority to remand a dispute back to a panel for further proceedings as a shortcoming. Others see the limited remedies available as a shortcoming, while still others are unwilling to accept a regime with stronger sanctions.

Nevertheless, since its inception the WTO dispute settlement system has been one of the most active forums, if not the most active, in the entire field of public international law. It has produced an extensive body of jurisprudence, and has, for the most part, been highly effective in its stated mission of "providing security and predictability to the multilateral trading system."

See also General Agreement on Tariffs and Trade (GATT); World Trade Organization

FURTHER READING

Hudec, Robert E. 1999. *Essays on the Nature of International Trade Law*. London: Cameron May. A superbly written, nontechnical collection of essays by the dispute settlement system's foremost commentator.

Jackson, John H. 1997. *The World Trading System: Law and Policy of International Economic Relations*. 2d ed. Cambridge, MA: MIT Press. An excellent overview of the workings of the rules governing the international trading system and the tensions those rules create.

Matsushita, Matsuo, Thomas J. Schoenbaum, and Petros C. Mavroidis. 2003. *The World Trade Organization: Law, Practice, and Policy*. Oxford: Oxford University Press. An analysis and commentary on the substantive provisions of the WTO agreements by three of the world's leading international trade law scholars.

Palmeter, David. 2003. *The WTO as a Legal System*. London: Cameron May. A collection of essays on legal and policy issues that have arisen in the WTO.

Palmeter, David, and Petros C. Mavroidis. 2004. *Dispute Settlement in the World Trade Organization: Practice and Procedure.* Cambridge: Cambridge University Press. A practice manual intended for diplomats and lawyers who participate in the WTO dispute settlement process.

Weiss, Friedl, ed. 2000. *Improving WTO Dispute Settlement Procedures: Issues and Lessons from the Practice of Other International Courts and Tribunals.* London: Cameron May. A collection of essays by leading academics looking at the WTO system in light of other international dispute settlement systems.

WTO Secretariat. 2001. *The WTO Dispute Settlement Procedures.* Cambridge: Cambridge University Press. The text of the WTO DSU, as well as the provisions from other WTO agreements relating to dispute settlement, and the texts of the formal decisions taken by the DSB with regard to practices and procedures.

———. 2004. *A Handbook on the WTO Dispute Settlement System.* Cambridge: Cambridge University Press. A general training guide to the system.

**NIALL P. MEAGHER
AND DAVID PALMETER**

Index

Note: Main entries are indicated by bold type.