Encyclopedia editor CARY L. COOPER
Advisory editors CHRIS ARGYRIS
AND WILLIAM H. STARBUCK

VOLUME VIII

THE BLACKWELL ENCYCLOPEDIA OF MANAGEMENT

SECOND EDITION

MANAGERIAL ECONOMICS

Edited by
ROBERT E. McAULIFFE

THE BLACKWELL ENCYCLOPEDIA OF MANAGEMENT

MANAGERIAL ECONOMICS

THE BLACKWELL ENCYCLOPEDIA OF MANAGEMENT

SECOND EDITION

Encyclopedia Editor: Cary L. Cooper
Advisory Editors: Chris Argyris and William H. Starbuck

THE BLACKWELL ENCYCLOPEDIA OF MANAGEMENT

SECOND EDITION

MANAGERIAL ECONOMICS

Edited by

Robert E. McAuliffe
Babson College

Contents

Preface

The second edition of the *Blackwell Encyclopedia of Management: Managerial Economics*, like the first edition, provides a reliable, comprehensive, and valuable resource for business practitioners, students, and researchers. The scope of the entries ranges from basic definitions such as the *law of demand* to advanced topics such as *estimating demand* and *time series forecasting models*. The entries are written clearly and concisely and often include references for those who wish to research the topics in more detail. The entries are also cross referenced so that readers may easily find information on related topics to obtain a thorough understanding of the concepts and their interrelationships.

The second edition of the encyclopedia also includes new references on the *economics of the Internet*, *network externalities*, and the *Microsoft antitrust case* (in both the US and the European Union). These entries help explain why the Internet was such an important development in business, why so many Internet firms failed (see *lock in*), and what the Internet has done to change government policy toward businesses.

As was the case in the first edition, this volume provides a careful exposition of the statistical and econometric issues that arise in applied work. Many editions of managerial economics texts include basic analysis of linear regression and statistical tests but frequently fail to mention the limitations of those tools. Entries in this volume on *simultaneous equations bias*, the *identification problem*, and *estimating demand* address these problems so that practitioners and researchers may avoid them in their work.

Robert E. McAuliffe

About the Editors

Contributors

Kostas Axarloglou
Athens Laboratory of Business Administration

Vickie L. Bajtelsmit
Colorado State University

Gilbert Becker
St. Anselm's College

Alexandra Bernasek
Colorado State University

Roberto Bonifaz
Adolfo Ibanez, Santiago, Chile

Michael D. Curley
Kennesaw State University

John Edmunds
Babson College

Govind Hariharan
Kennesaw State University

Sean Hopkins
Renaldo and Palumbo Law Firm, Clarence, NY

Sayed Ajaz Hussain
Babson College

Kent A. Jones
Babson College

Brett Katzman
Kennesaw State University

Eduardo Ley
International Monetary Fund

Wei Li
Darden Business School, University of Virginia

Robert E. McAuliffe
Babson College

Alastair McFarlane
Department of Housing and Urban Development

Steven G. Medema
University of Colorado at Denver

Dileep R. Mehta
Georgia State University

Maria Minitti
Babson College

Laurence S. Moss
Babson College

Lidija Polutnik
Babson College

Laura Power
Treasury Department, Washington, DC

Mark Rider
Georgia State University

Don Sabbarese
Kennesaw State University

S. Alan Schlacht
Kennesaw State University

James G. Tompkins
Kennesaw State University

Theofanis Tsoulouhas
North Carolina State University

Roger Tutterow
Kennesaw State University

Nikolaos Vettas
Athens University of Economics and
Business

A

accommodation

Robert E. McAuliffe

When established firms are threatened by entry, those firms can retaliate against the new firm by cutting prices and increasing advertising, or the firms can accommodate the entrant. Entry will be accommodated when the incumbent firms cannot maximize profits by deterring entry and they will not react aggressively to the entrant because it would lower their profits to do so. The incumbent firms may have FIRST MOVER ADVANTAGES since they can take actions before entry which may affect the entrant's profitability and market position. For example, existing firms may under or overinvest in capital equipment to influence the entrant's choice of its scale of operation, even though the existing firms cannot prevent the entry (see Jacquemin, 1987; Tirole, 1988).

Whether or not the existing firms over or underinvest in capital depends on how the investment will affect the existing firms' competitive position in the period(s) after entry has occurred and on the expected reaction of the entering firm if entry were to occur. For example, investment in productive capacity or in producing additional output to experience lower costs from the LEARNING CURVE all make the incumbent firms tougher competitors in the second period. If entry can be deterred, the established firms will then overinvest in the current period. But if the incumbents cannot deter the entrant and the entrant is expected to react aggressively, they should underinvest in these activities to avoid the aggressive response from the entrant in the next period. ADVERTISING, however, can make incumbent firms less likely to respond aggressively to entry since they can enjoy higher profits from the "captive" consumers who have received advertising messages (Schmalensee, 1983). Therefore in some cases established firms should underinvest in advertising to indicate their willingness to aggressively compete against an entrant should entry occur.

Bibliography

Fudenberg, D. and Tirole, J. (1984). The fat-cat effect, the puppy-dog ploy and the lean and hungry look. *American Economic Review*, **74**, 361 6.

Jacquemin, A. (1987). *The New Industrial Organization: Market Forces and Strategic Behavior*. Cambridge, MA: MIT Press.

Schmalensee, R. (1983). Advertising and entry deterrence: An exploratory model. *Journal of Political Economy*, **91**, 636 53.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

accounting profit

Robert E. McAuliffe

Accounting profit is defined as total revenues from output sold in a given period minus those costs incurred during that period (including DEPRECIATION expenses). The difference between the accounting definition of profits and economic profit lies in how costs and depreciation are calculated (*see* ECONOMIC PROFIT). Under generally accepted accounting practices, all costs incurred by the firm in a given period are expensed in that period (except expenditures on tangible assets, which are depreciated over several periods). This means that expenditures on research and development, training, trade marks, goodwill, and patents – all sources of intangible capital – are expensed in the current period, even though they may yield benefits well

into the future. As a result, accounting profits will overstate economic profits whenever current period profits were generated in part by previous investments in intangible assets because there are no accounting costs applied in the current period for those intangible assets and those assets are not often included in calculations of the firm's total value. In addition, the depreciation expense for tangible assets allowed under accounting rules is not the same as the ECONOMIC DEPRECIATION for those assets.

These problems with accounting measures of economic profits have led some economists to argue that there is no relationship between accounting profits and economic profits (Fisher and McGowan, 1983). This strong assertion has been challenged by several economists and remains controversial (see Long and Ravenscraft, 1984; Martin, 1984). Salamon (1985) and Edwards, Kay, and Mayer (1987) provide recommendations regarding proper adjustments of accounting profits and those circumstances where they will more reliably approximate economic profits. Salamon suggests using conditional INTERNAL RATE OF RETURN (IRR) estimates from financial statements as a proxy for the economic rate of return which can be used to infer the measurement errors from using accounting profits. He found that accounting rates of return, while strongly correlated with the estimated IRR, nevertheless showed considerable variation that the IRR could not explain. The measurement error from using accounting rates of return was systematically related to firm size and therefore cast doubts on cross-section studies of the relationship between concentration and profitability.

Bibliography

Edwards, J., Kay, J., and Mayer, C. (1987). *The Economic Analysis of Accounting Profitability*. Oxford: Oxford University Press.
Fisher, F. M. and McGowan, J. J. (1983). On the misuse of accounting rates of return to infer monopoly profits. *American Economic Review*, **73**, 82–97.
Long, W. F. and Ravenscraft, D. J. (1984). The misuse of accounting rates of return: Comment. *American Economic Review*, **74**, 494–500.
Martin, S. (1984). The misuse of accounting rates of return: Comment. *American Economic Review*, **74**, 501–6.
Salamon, G. L. (1985). Accounting rates of return. *American Economic Review*, **75**, 495–504.

adverse selection

Alexandra Bernasek

Adverse selection refers to a class of problems that are created by ASYMMETRIC INFORMATION between parties to a transaction. Adverse selection problems arise because the characteristics of products sold in markets differ, and one party (generally the seller) has valuable information about those characteristics which is not available to the other party or parties (generally the buyers). The classic illustration of this kind of problem is Akerlof's (1970) modeling of the used car market (*see* LEMONS MARKET). Adverse selection problems are also a consequence of individuals having different abilities, and there being IMPERFECT INFORMATION about a specific individual's abilities. In those cases one party to a transaction has valuable information about their own ability, but that information is not available to the other party or parties. Adverse selection problems often arise in the context of principal–agent problems (*see* PRINCIPAL–AGENT PROBLEM). They are common in insurance markets, financial markets, and labor markets.

Adverse selection has consequences for market EFFICIENCY. In the presence of adverse selection, the allocation of resources is almost always inefficient, and under certain conditions EQUILIBRIUM may not exist. Rothschild and Stiglitz (1976) establish these effects of adverse selection in insurance markets. The adverse selection problem in insurance markets arises because those with the highest probability of experiencing a negative event are the ones who want to purchase insurance, but they are the least desirable customers from the perspective of the insurance company because of their high probability of becoming claimants. The insurance company problem is that it has difficulty distinguishing between the different types of individuals.

Stiglitz and Weiss (1981) examine the effects of adverse selection in financial markets. Adverse selection problems arise in credit markets

because it is difficult for lenders to distinguish between individuals who have a high probability of default and those who have a low prob ability. Stiglitz and Weiss find that when banks employ screening devices such as raising interest rates or collateral requirements, these can affect the behavior of borrowers and the distribution of borrowers (for example those who are willing to borrow at high interest rates may be worse risks on average) and can increase the riskiness of the bank's portfolio; thus banks may be more likely to ration credit instead. Their results have im plications for landlord–tenant relationships and employer–employee relationships as well.

A classic work on how agents try to overcome problems of adverse selection is Spence's (1974) work on market SIGNALING. He suggests that the more able (higher quality) individuals will want to signal their ability to the other parties in the transaction. In the context of the labor market, for example, those individuals may be willing to incur costly education or training in order to signal their quality to an employer. In the context of the used car market on the other hand, sellers of the "good" cars may be willing to incur the cost of offering a warrantee with the sale of the car.

See also *asymmetric information; imperfect in formation; principal–agent problem*

Bibliography

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Jour nal of Economics*, **84**, 488 500.

Rothschild, M. and Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the eco nomics of imperfect information. *Quarterly Journal of Economics*, **90**, 629 49.

Spence, M. (1974). *Market Signaling*. Cambridge, MA: Harvard University Press.

Stiglitz, J. and Weiss, A. (1981). Credit rationing in markets with imperfect information. *American Eco nomic Review*, **71**, 912 27.

advertising

Robert E. McAuliffe

Advertising refers to expenditures in various media (such as radio, television, newspapers, magazines, etc.) made by firms to increase sales. Firms may often use advertising to differ entiate their products from competing brands. If successful, advertising could increase the demand for a product, reduce the price ELAS TICITY of demand, and allow the firm to charge a higher price and earn higher profits. Advertis ing has been the subject of some controversy in terms of both how it affects the demand for a firm's product and if unregulated markets gen erate too little or too much advertising. One important issue is whether advertising increases sales by changing consumer tastes (persuasion) or by informing consumers of alternative brands. If advertising increases sales merely by persuad ing consumers, then society is not necessarily better off as a result of advertising expenditures since these outlays have arbitrarily altered tastes in favor of the advertised product. But if adver tising provides information, these expenditures may enable consumers to make better choices in the market and reduce their costs of finding appropriate products. When advertising pro vides information to consumers the expend itures may increase EFFICIENCY.

OPTIMAL ADVERTISING LEVELS

The profit maximizing level of advertising occurs when the MARGINAL REVENUE from additional advertising expenditures is just equal to the MARGINAL COST. If advertising allows the firm to sell one more unit, then the marginal benefit from the sale is the firm's profit per unit, price minus the marginal cost of production – or $(P - MC)$ – and this should be equated to the marginal cost of advertising. The additional sales revenue a firm can expect from advertising will depend upon the number of potential consumers exposed to the advertisement and that advertise ment's effectiveness in creating a sale. The mar ginal cost of advertising should be rising as more advertising messages are sent to consumers since additional advertising messages will have de creasing effectiveness as the number of mes sages increases (*see* the DORFMAN–STEINER CONDITION).

To maximize profits from a given advertising budget, firms should advertise in different media until the marginal profit per dollar spent on each of the media is equal. If a firm earns a profit of $(P - MC)$ per unit sold, and P_{TV} is the price of

an advertising message on television, the mar
ginal profit per dollar spent on television is
$(P − MC)/P_{TV}$ for every additional sale
brought about by a television advertisement.
Similarly, the marginal profit per dollar
spent on newspaper advertisements will be
$(P − MC)/P_{NEWS}$ and the firm should advertise
until it earns the same expected net profits in
each of the media. That is:

$$ES_{TV}(P − MC)/P_{TV}$$
$$= ES_{NEWS}(P − MC)/P_{NEWS}$$
$$= \ldots$$
$$= ES_{RADIO}(P − MC)/P_{RADIO}$$

where ES_i represents the expected additional
units sold from advertising messages in each of
the i media, P_i is the price of an advertising
message in each of the media, and P_{RADIO} is
the price of an advertising message on radio.
Following this principle, the firm will maximize
expected profits from its advertising budget.

Advertising intensity has often been measured
by the advertising–sales ratio, that is, total ad
vertising expenditures divided by total sales.
While the Dorfman–Steiner condition shows
that the profit maximizing level of advertising
depends on this ratio, it may not be an appropri
ate measure of advertising intensity. Consumers
do not respond to the dollar amount a company
spends on advertising, they respond to the
number of messages they see. Therefore the
appropriate measure of advertising intensity for
managers should be total advertising expend
itures deflated by an index of the cost per million
viewers in that medium. This adjusted measure
will indicate the number of people potentially
exposed to an advertising message and could be
divided by sales to measure advertising intensity
(see Ehrlich and Fisher, 1982; McAuliffe, 1987).

DOES ADVERTISING INCREASE OR REDUCE COMPETITION?

If advertising creates a barrier to entry (*see* BAR
RIERS TO ENTRY), established firms could
enjoy long run economic profits (*see* ECO
NOMIC PROFIT). Advertising could create
brand loyalty, for example, and decrease the
price elasticity of demand. According to Coma

nor and Wilson (1974), advertising expenditures
could create a barrier to entry by increasing the
capital required for entry, creating ECONOMIES
OF SCALE, or creating brand loyalty for estab
lished brands. If a new entrant has to advertise
more to overcome brand loyalty, the entrant
could be placed at a disadvantage relative to
existing firms. Simon and Arndt (1983) pointed
out that it is incorrect to argue that advertising
(or any other single input) can create economies
of scale because the concept refers to changes in
all inputs. When costs decline as more of a *single*
input is employed, there are increasing returns
to that input. But Simon and Arndt found there
are DIMINISHING RETURNS to advertising.
Spence (1980) developed a model where adver
tising is treated as an input in the production of
sales revenue for the firm and suggested that
advertising could combine with other factors of
production to create economies of scale advan
tages for established firms. He showed that es
tablished firms could use these economies of
scale to their advantage to deter entry.

Since established firms are already in the
market, Cubbin (1981) argued they could
have FIRST MOVER ADVANTAGES. Incumbent
firms could use this strategic advantage by in
creasing their advertising so that potential en
trants would have to advertise more as well. But
as Schmalensee (1983) and Fudenberg and Tir
ole (1984) have indicated, advertising to prevent
entry, much like LIMIT PRICING, is a reversible
decision. The threat to increase advertising or
output may not be the profit maximizing choice
once entry has occurred, since established firms
may earn higher profits if they accommodate the
entrant. In such a case, the threat of higher
advertising or higher output is not a credible
strategy (*see* CREDIBLE STRATEGIES). When
Schmalensee examined the post entry equilib
rium, he found that established firms always
advertised less when threatened by entry. This
unusual result occurs if the established firm has
become a "fat cat," to use Fudenberg and Tir
ole's terminology. When advertising creates
goodwill it will have two effects on potential
entry. First, advertising by the incumbent will
reduce the market share that remains free to the
potential entrant and this reduces the incentive
to enter the industry. But as this goodwill

increases, the established firm becomes fat and lazy. Since the established firm has a loyal customer base from its past advertising, it is less inclined to react aggressively to a new entrant and will not increase its advertising expenditures or reduce prices. When the latter effect is stronger, the established firm must *underinvest* in advertising to signal to potential entrants that it will aggressively cut prices and increase advertising if entry were to occur. What is important from this literature is that even if established firms have the ability to prevent entry, that does not mean that it is in their best interests to do so. Profits may be higher if the rate of entry is reduced (*see* ACCOMMODATION; CREDIBLE STRATEGIES; SIGNALING).

Nelson (1974) suggested that the elasticity of demand for a product depended on the number of *known* alternative brands. Since advertising provides information about the existence of competing products, he argued that advertising could increase the elasticity of demand and make entry into an industry less difficult. Furthermore, Nelson argued that the information content of advertising and the best media to choose would differ depending upon whether the product was a search good or an experience good (*see* SEARCH GOODS; EXPERIENCE GOODS). Since consumers can easily determine the characteristics of search goods by inspection, advertising for search goods will tend to be informative and concentrated in more informative media such as newspapers and magazines. But consumers must actually use experience goods to determine their quality, so informational advertising will not be as helpful. Therefore advertising for experience goods relies more on product imagery and seller reputation while being concentrated in more experiential media such as television. Furthermore, Nelson showed that even if advertising for experience goods was not informative, the very fact that the product was advertised conveyed information to consumers about the product's quality. Nelson suggested that only high quality producers would have the incentive to advertise heavily and this served as a signal of quality to consumers of the good.

Ehrlich and Fisher (1982) agreed that advertising provided information and that it reduced the costs to consumers of finding those products which best fulfilled their needs. They also distinguished between media advertising expenditures and other promotional efforts. Since advertising provides information, both firms and consumers can produce the information, though at different costs. The full price of a product to the consumer is the price paid plus information costs, which Ehrlich and Fisher assert are primarily the consumer's time searching for the appropriate product. Firms can reduce these costs through media advertising expenditures or through other promotional efforts such as trade shows, customer services, and other selling efforts. They predicted advertising and selling efforts would be greater the higher are consumer wages and the larger the market for the brand. Higher wages imply higher time costs to consumers from search while a larger market implies a lower cost to the manufacturer of providing information through advertising. Furthermore, media advertising should be less for producer goods since these buyers are very knowledgeable and identifiable relative to buyers of consumer goods. Therefore they predicted that trade shows and other direct selling methods would be more productive for these products. The empirical results using US data from 1946 to 1969 support their hypotheses: advertising and promotional efforts are positively related to wage rates in the economy and advertising–sales ratios are negatively related to the price elasticity of demand. Ehrlich and Fisher also found that advertising expenditures did not have long lived effects and were completely depreciated within one year, a finding that is also consistent with McAuliffe (1987).

McAuliffe (1987) tested the hypothesis that advertising reduced competition and increased firm profitability. If advertising causes higher profits, then current advertising levels should have significantly positive effects on future profits. Out of 27 firms for which there were data from 1955 to 1983, advertising had significant, consistently positive effects on future profits for only three firms. While there was strong correlation between current advertising and current profits for the firms in the sample, the effects of advertising did not last beyond a year.

There is an important difference between advertising expenditures and other capital ex penditures from the entrant's perspective, how ever. If the attempt to enter the industry does not succeed, the entrant can recover some of its original investment costs by selling its plant and equipment. But advertising costs cannot be re covered and thus represent a sunk cost to the entering firm (*see* SUNK COSTS). When entry requires significant levels of advertising, it is more risky, the costs of failing are that much higher, and this could deter entrants (see Kes sides, 1986; Sutton, 1991). In his study, Sutton suggested that advertising and research and de velopment represent *endogenous* sunk costs be cause firms can vary the amounts of these expenditures, while the sunk costs of investment in plant and equipment is dictated by technol ogy. This means that in some industries firms may engage in ever escalating expenditures in these areas as they compete to gain advantage.

Bibliography

Comanor, W. S. and Wilson, T. A. (1974). *Advertising and Market Power*. Cambridge, MA: Harvard University Press.

Cubbin, J. (1981). Advertising and the theory of entry barriers. *Economica*, **48**, 289 99.

Ehrlich, I. and Fisher, L. (1982). The derived demand for advertising: A theoretical and empirical investigation. *American Economic Review*, **72**, 366 88.

Fudenberg, D. and Tirole, J. (1984). The fat-cat effect, the puppy-dog ploy and the lean and hungry look. *American Economic Review*, **74**, 361 6.

Kessides, I. N. (1986). Advertising, sunk costs, and bar riers to entry. *Review of Economics and Statistics*, **68**, 84 95.

McAuliffe, R. E. (1987). *Advertising, Competition, and Public Policy: Theories and New Evidence*. Lexington, MA: D. C. Heath.

Nelson, P. (1974). Advertising as information. *Journal of Political Economy*, **82**, 729 54.

Schmalensee, R. (1983). Advertising and entry deter rence: An exploratory model. *Journal of Political Econ omy*, **91**, 636 53.

Simon, J. L. and Arndt, J. (1983). Advertising and econ omies of scale: Critical comments on the evidence. *Journal of Industrial Economics*, **32**, 229 42.

Spence, M. A. (1980). Notes on advertising, economies of scale, and entry barriers. *Quarterly Journal of Econom ics*, **95**, 493 507.

Sutton, J. (1991). *Sunk Costs and Market Structure*. Cam bridge, MA: MIT Press.

antitrust policy (US)

Gilbert Becker

The last half of the nineteenth century intro duced a period of rapid industrial growth in the US. According to Brock (see Adams and Brock, 2001) and others, Congress, which a century earlier had been focused on gaining independ ence from growing British political power over the US colonies, now became concerned with the amassing of economic power in the hands of a few. It chose not to nationalize these de veloping oligopolies (*see* OLIGOPOLY), and in only a few instances turned to regulatory com missions (*see* NATURAL MONOPOLY), which tend to replace basic market functions such as pricing and ENTRY decisions with their own process. Instead, Congress opted to design a set of antitrust laws which could promote the dual overriding goals of (1) retaining and supporting CAPITALISM and its economic freedoms, while (2) controlling monopoly and fostering a more competitive environment. The antitrust ap proach enabled the continuance of a central tenet of capitalism, that of privately held eco nomic power, while at the same time helped to disperse that power among many. Thus, anti trust policy was an effort to "substitute the decentralized decision making system of the competitive market for central planning, whether by the state or, alternatively, by private monopolists, oligopolists, or cartels" (Adams and Brock, 2001: 366).

Antitrust policy in the US and European Union (EU) is based on several statutes which identify various forms of business behavior that may be anticompetitive and therefore illegal. Managers should be aware that these business practices include price fixing, mergers, PREDA TORY PRICING, PRICE DISCRIMINATION, TYING, attempting to create a MONOPOLY, developing and maintaining CARTELS, and other conduct. The three main antitrust statutes in the US are the 1890 SHERMAN ACT, the 1914 CLAYTON ACT, and the 1914 FEDERAL TRADE COMMISSION ACT. There are coun terparts concerning competition policy in the EU which are found in Articles 85 and 86 of the 1957 Treaty of Rome (*see* EU COMPETI TION POLICY, 2004). The laws of these two regions share certain broad conceptual

foundations yet retain differences in their appli
cation, enforcement, and to a certain extent their
underlying objectives. Each of these laws has
several sections that are written in language
which is open to interpretation. As a result, the
implementation of policy is dependent (in part)
on the philosophies of the members of the gov
ernmental agencies and judiciary who are em
powered, at any given point in time, to enforce
the law.

Debate in the US has continued for decades as
to the original intent of the framers of the Sher
man and Clayton Acts. Bork (1966) believes that
Senator Sherman was concerned with the reduc
tion of output and DEADWEIGHT LOSS ineffi
ciency resulting from monopoly. Martin (1994)
and others believe that their original intent was
also to protect consumers from unfair prices
yielding excess economic profits (see ECO
NOMIC PROFIT) and in some instances to pro
tect small businesses from unfair practices
of their larger rivals. Still others, including
Katzman (1984), argue that the original intent,
in part, grew out of the concern that economic
power, from both large absolute size and large
relative size of firms, may translate into political
power, to the detriment of democracy and the
country's social structure. The interpretation of
these statutes and the rigor with which they are
enforced have varied over time and likely will
continue to do so. In a historical analysis of the
first century of antitrust in the US, Schwartz
(1990) details cycles of approximately 25 years
between peak periods of aggressive antitrust en
forcement.

There is also controversy concerning the
economic theory supporting these statutes
(see STRUCTURE CONDUCT PERFORMANCE
PARADIGM). Kovaleff (1990) has compiled the
works of numerous current antitrust scholars
who provide an array of studies on the merits
of US antitrust law at the conclusion of its first
century. Much of this work continues to focus
on the perceived trade off between abuses of
economic power and the benefits of economic
EFFICIENCY resulting from large firm size. In
addition, new economic concepts such as NET
WORK EXTERNALITIES have recently brought
into question the application of century old anti
trust laws to modern industrial markets (see
MICROSOFT ANTITRUST CASE).

Bibliography

Adams, W. and Brock, J. (2001). *The Structure of Ameri
can Industry*, 10th edn. Upper Saddle River, NJ:
Prentice-Hall.

Bork, R. (1966). Legislative intent and the policy of the
Sherman Act. *Journal of Law and Economics*, **9**, 7 48.

Breit, W. and Elzinga, K. (1996). *The Antitrust Casebook:
Milestones in Economic Regulation*, 3rd edn. Fort Worth,
TX: Dryden Press.

Katzman, R. (1984). The attenuation of antitrust. *The
Brookings Review*, **2**, 23 7.

Kovaleff, T. (1990). A symposium on the 100th anniver
sary of the Sherman Act. *Antitrust Bulletin*, **35**.

Martin, S. (1994). *Industrial Economics*, 2nd edn. New
York: Macmillan.

Schwartz, L. (1990). Cycles of antitrust zeal: Predictabil
ity? *Antitrust Bulletin*, **35**, 771 800.

Shepherd, W. and Shepherd, J. (2004). *The Economics of
Industrial Organization*, 5th edn. Long Grove: Wave
land Press.

antitrust remedies (US)

Gilbert Becker

There are numerous methods by which the US
antitrust laws allow for remedial action or penal
ties to be levied subsequent to a finding that the
law has been violated. SHERMAN ACT viola
tions are criminal felonies. Thus, where the law
is clear and the intent to violate it can be inferred
by the existence of certain conduct (e.g., price
fixing), criminal penalties may be imposed.
These include corporate fines of up to $100
million, individual fines of up to $10 million,
and prison sentences for individuals of up to
three years. A recent example is that of the
former Chairman of Sotheby's who, in 2002,
was sentenced to one year in prison and a fine
of $7.5 million for his guilt in a price fixing
conspiracy (see Markon, 2002).

Where the law is less clear, civil action cases
may be brought wherein remedial action without
penalties may result. This is the case for most
Sherman Act and all CLAYTON ACT and FED
ERAL TRADE COMMISSION ACT violations.
These remedial actions include equitable relief,
which can be achieved through the restraint of

anticompetitive conduct or the cancellation of contracts or even the divestiture of certain assets. Remedy may also arise from a consent decree in which a corporation admits no guilt but agrees to relief which is acceptable to the government and sanctioned by the courts. With respect to the Federal Trade Commission Act, the Commission holds additional remedial powers including the ability to issue cease and desist orders to violators of Section 5 of the Act.

As Howard (1983) points out, defendants in antitrust cases brought by the government may choose to plead *nolo contendere*, not contending the charge(s) brought against them. This strategy, which does not require admission of any wrongdoing, results in no prima facie evidence being established against the firm. Here, the court may administer whatever remedy it sees as being necessary and appropriate. One possible benefit to the firm from such a plea, along with avoiding lengthy and costly court proceedings, concerns additional private antitrust suits which may be brought against the firm by injured parties. As a result of the plea, injured parties seeking damages would have to develop their own evidence establishing that a violation of the law had in fact occurred. The cost of the case development may deter some injured parties from initiating action.

In monopolization cases, a basic decision concerning remedies is whether the relief should be *structural* or *behavioral* in nature. Structural relief follows the economic theory that MARKET STRUCTURE often generates certain forms of conduct and may ultimately result in undesirable MARKET PERFORMANCE. It calls for some form of breaking up of the monopolist in an effort to return the market to a more competitive state (*see* STRUCTURE CONDUCT PERFORMANCE PARADIGM). This may include the dissolution of a company into several smaller rivals, as was the case in US v. *Standard Oil Company of New Jersey* in 1914, or the divestiture of some assets and the creation of separate firms, as was the result of the court approved settlement between the government and the AT&T Corporation in 1982.

A less harsh behavioral approach would focus on the firm's conduct and prevent the firm from taking certain anticompetitive actions. This approach may also be used in conjunction with structural relief (see MICROSOFT ANTITRUST CASE: REMEDIES).

Antitrust law also allows for relief to be reached by private lawsuits for monetary damages. Section 4 of the Sherman Act allows injured parties to recover sums which are three times the amount of the monetary damages incurred. In addition, the antitrust violator must compensate the injured party for all reasonable attorney's fees and other legal expenses resulting from the lawsuit.

The vast majority of antitrust cases filed in court are private suits, perhaps in large part because of these treble damage awards. The benefits and potential abuses of the treble damages system, and the incentives of the antitrust law penalty system in general, have come under heightened scrutiny by Breit and Elzinga (1986), Werden and Simon (1987), Grippando (1989), and others. Moreover, the business strategies of several firms today (especially those with rapidly changing technology) have enlivened the debate surrounding antitrust policy in the US and other nations. Boudreaux and Folsom (1999) argue that antitrust action inevitably harms consumers in markets with rapid technological development, in part by dampening entrepreneurial creativity (see Schumpeter, 1950), but also because the specialized knowledge necessary for entrepreneurial decision making today leaves antitrust enforcers incapable of judging business decisions in such a way as to improve on market generated outcomes. Flynn (2001) disagrees, finding that the anticompetitive behavior of Microsoft holds "intriguing parallels" to those of Standard Oil 90 years earlier, and as such he sees the Sherman Act equally applicable today.

Finally, the traditional antitrust remedies have come under attack due to the recent development of economic concepts such as NETWORK EXTERNALITIES. As both the antitrust laws and the success of these industries are important to the economy, the evolution toward any new public policy outcome is likely to be gradual.

See also *antitrust policy (US); EU competition policy, 2004; Microsoft antitrust case*

Bibliography

Boudreaux, D. and Folsom, B. (1999). Microsoft and Standard Oil: Radical lessons for antitrust reform. *Antitrust Bulletin*, **44**, 555 77.

Breit, W. and Elzinga, K. (1986). *Antitrust Penalty Reform*. Washington, DC: American Enterprise Institute.

Flynn, J. (2001). Standard Oil and Microsoft: Intriguing parallels. *Antitrust Bulletin*, **46**, 4, 645 734.

Grippando, J. (1989). Caught in the non-act: Expanding criminal antitrust liability for corporate officials. *Anti trust Bulletin*, **34**, 713 57.

Howard, M. (1983). *Antitrust and Trade Regulation*. Englewood Cliffs, NJ: Prentice-Hall.

Jorde, T., Lemley, M., and Mnookin, R. (1996). *Gilbert Law Summaries: Antitrust*, 9th edn. Chicago: Harcourt Brace Legal and Professional Publications.

Markon, J. (April 23, 2002). Ex-Sotheby's chairman is sentenced to year in prison, fined $7.5 million. *Wall Street Journal*, **239**, 79, B4.

Schumpeter, J. (1950). *Capitalism, Socialism, and Democracy*, 3rd edn. New York: Harper and Row.

Waldman, D. and Jensen, E. (2001). *Industrial Organization: Theory and Practice*, 2nd edn. Boston: Addison-Wesley.

Werden, G. and Simon, M. (1987). Why price fixers should go to jail. *Antitrust Bulletin*, **32**, 913 37.

arbitrage

Robert E. McAuliffe

Arbitrage is the process of buying goods or assets in one market where the price is lower and selling them in markets where the price is higher for riskless profits. To successfully practice PRICE DISCRIMINATION firms must prevent arbitrage from occurring between the markets in which they sell the product at different prices. The act of arbitrage will tend to equalize prices between the two markets as demand is increased in the market where the price is lower, causing its price to rise, and supply is increased in the market where the price is higher, causing its price to fall. This process may profitably con tinue until the difference in price between the two markets is equal to the transportation costs of moving the good from one market to the other.

Since arbitrage occurs frequently and easily in financial markets, modern finance theory relies on arbitrage arguments to understand asset pricing (see Varian, 1987).

Bibliography

Varian, H. R. (1987). The arbitrage principle in financial economics. *Journal of Economic Perspectives*, **1** (2), 55 72.

arc elasticity

Gilbert Becker

Arc elasticity is the measure of ELASTICITY to be used when the effect of a large change in a variable (e.g., price) is examined. Price elasticity of demand, ε_p, defined as the percentage change in quantity demanded for a given percentage change in price, can be calculated as

$$\varepsilon_p = \frac{(Q_2 - Q_1)}{(P_2 - P_1)} \times \frac{P}{Q}$$

where the subscripts indicate initial (1) and final (2) values for price (P) and quantity (Q). When large price changes are used, the value of the second term (P/Q), and thus of ε_p, may vary sharply depending on whether the initial or final price and quantity values are used. Arc elasticity solves this problem by using the aver age price and quantity over the ranges in ques tion. This gives an approximation of the consumer responsiveness for the entire range. As business pricing strategies typically involve discrete changes (for example, a 10 percent off sale), arc elasticity is often the appropriate meas ure for management to examine.

See also *elasticity*

Bibliography

Browning, E. and Zupan, M. (2004). *Microeconomics: Theory and Applications*, 8th edn. New York: John Wiley.

Maurice, S. and Thomas, C. (2002). *Managerial Economics*, 7th edn. Boston: McGraw-Hill Irwin.

Miller, R. L. (2004). *Economics Today: The Micro View*, 12th edn. Boston: Pearson.

asset specificity

Robert E. McAuliffe

An asset is specific if it has high value only when used in certain applications and does not have much value in alternative uses. Asset specificity is also an attribute of a given transaction and since it represents a greater risk to one party in the transaction, the costs of that transaction will be higher. This can create problems in contracting between firms in cases where, say, a supplier might have to make investments that are specific to its customer. The problem is that such an investment leaves the supplier vulnerable to the whims of its customer and the customer could put the supplier at a disadvantage – what the transaction cost literature refers to as the "hold up problem" (Milgrom and Roberts, 1992; *see* TRANSACTIONS COSTS). Both parties have an interest in resolving this problem, and economic incentives suggest that they will in those cases where transactions costs are not too high. But it is possible that the supplier might avoid making the necessary investments in specific assets if the uncertainty is too great, and asset specificity is one element that raises transactions costs. All else equal, a transaction will be more easily undertaken when both parties have little to risk. To limit its risk, a supplier might require complex, long term CONTRACTS with its buyer to safeguard its investments in specific assets (see Joskow, 1987), or the two firms may integrate vertically to internalize these transactions costs and avoid the hold up problem. In fact, Williamson (1986) has argued that asset specificity provides the major motivation for VERTICAL INTEGRATION.

Williamson identifies four different kinds of asset specificity which affect the decision to organize activities within the firm versus through the market.

1 Site specificity: when an asset such as a plant must be located at a particular site to meet the requirements of the buyer. This can arise when, for example, railroads provide service to deliver coal to an electric utility. The track investment is not valuable for any other customers other than the utility.
2 Physical asset specificity: if the rail cars needed to transport the coal to the utility are unique and have little value outside of that purpose, then the railroad's investment in these cars would represent a specific physical asset.
3 Human asset specificity: when people acquire skills specific to their work at the firm or in particular teams, their skills may not be as valuable in any other firms or with other teams. In these cases, an employment arrangement rather than a market arrangement would be the expected form of organization because workers are less likely to invest in acquiring skills that are valuable (specific) to only one firm.
4 Dedicated assets: if a producer must expand capacity to meet the needs of a buyer, that producer now bears more risk and may require contractual assurances from the buyer.

These transactions costs affect the optimal size of the firm, the MINIMUM EFFICIENT SCALE. As Coase (1937) suggested, there are costs to using the market just as there are costs to organizing activities internally within the firm. Decisions made within the firm are made by HIERARCHY and can be less costly than relying on the market. The boundary of the firm is determined at the point where it is less costly to use the market to obtain goods and services than to produce them within the firm. As Williamson (1986) notes, this is essentially a make or buy decision for the firm (*see* MAKE OR BUY DECISIONS). Transactions costs are higher from using the market when contracts are difficult to write that will prevent one party from taking advantage of the other or when contracts are incomplete. A well known example discussed by Klein, Crawford, and Alchian (1978) is the arrangement between GM and the Fisher Body plant. Fisher was unwilling to make the investments in specific assets required by GM because the plant would be of little use to any company other than GM. Ultimately, GM integrated backward and purchased Fisher Body.

Bibliography

Coase, R. H. (1937). The nature of the firm. *Economica*, **4**, 386 405.

Joskow, P. (1987). Contract duration and durable transaction-specific investments: The case of coal. *American Economic Review*, **77**, 168 85.

Klein, B., Crawford, R., and Alchian, A. (1978). Vertical integration, appropriable rents, and the competitive contracting process. *Journal of Law and Economics*, **21**, 297 326.

Milgrom, P. and Roberts, J. (1992). *Economics, Organiza tion and Management*. Englewood Cliffs, NJ: Prentice-Hall.

Williamson, O. E. (1986). Vertical integration and related variations on a transactions-cost economics theme. In J. E. Stiglitz and G. F. Mathewson (eds.), *New Develop ments in the Analysis of Market Structure*. Cambridge, MA: MIT Press.

asymmetric information

Robert E. McAuliffe

Asymmetric information exists when one party in the market or transaction has more or better information than the other party. Furthermore, the party with less information cannot rely on the other for the necessary information and cannot easily acquire it. For example, Akerlof (1970) noted that sellers of used cars have more information about the quality of the used car than buyers. This asymmetry can place the party with less information at a disadvantage and can interfere with market exchange to the point where market transactions break down. Akerlof showed that if buyers in the used car market considered all cars to be "average" in quality, no sellers of above average quality cars would want to sell. This would reduce the average quality of the cars remaining in the market until only the worst cars (lemons) were traded (*see* ADVERSE SELECTION; IM PERFECT INFORMATION; LEMONS MARKET). There is a tendency in these markets for qual ity levels to fall if consumers cannot discrimin ate between high quality and low quality products. The problem of asymmetric infor mation also arises in employment decisions, insurance markets, and credit markets where the person who is applying for a job, insurance, or credit knows more about his or her abilities, health, or risk than the employer, insurer, or creditor. Firms have incentives to acquire more information in these situations while job applicants and consumers have incentives to provide more information, perhaps through SIGNALING.

High quality producers have incentives for signaling in these markets to convince consumers that their products are better than average. Guar antees or warranties can be provided to assure consumers that a product will perform above the average. Firms also have incentives to invest in their reputation and in brand names to indicate that the product is a high quality product. The product's price itself may convey information about quality in the appropriate circumstances. For example, Milgrom and Roberts (1986) found that both price and ADVERTISING could provide signals to consumers for new EXPERIENCE GOODS that are frequently purchased. High quality firms have incentives to set a low price because they will benefit more from future repeat purchases than low quality producers. However, Bagwell and Riordan (1991) suggest that for a new durable product the initial price should be high to signal that it is a high quality product to uninformed consumers. As sales occur and more of the market becomes informed that this is a high quality product, the firm should decrease price to maximize profits.

Bibliography

Akerlof, G. A. (1970). The market for lemons: Quality uncertainty and the market mechanism. *Quarterly Jour nal of Economics*, **84**, 488 500.

Bagwell, K. and Riordan, M. K. (1991). High and declin ing prices signal product quality. *American Economic Review*, **81**, 224 39.

Milgrom, P. and Roberts, J. (1986). Price and advertising signals of product quality. *Journal of Political Economy*, **94**, 796 821.

Shugart, W. F., Chappell, W. F., and Cottle, R. L. (1994). *Modern Managerial Economics*. Cincinnati, OH: South-Western Publishing.

Stiglitz, J. E. (1987). The causes and consequences of the dependence of quality on price. *Journal of Economic Literature*, **25**, 1 48.

auctions

Brett Katzman

Auctions are institutions used to sell products or award contracts (such as government procure ment). They are commonly employed when there is no established market for the good and thus no existing price. A single seller of the

product to be auctioned has some degree of MARKET POWER as a monopolist or as the buyer of services as a monopsonist (*see* MONOPOLY; MONOPSONY). However, at sites such as eBay, this market power is often diluted by a multiplicity of sellers.

There are four primary types of auctions. Perhaps the best known is the **English** (or oral ascending) **auction** where buyers compete by signaling (usually done with oral outcries) increasingly higher bids until no further bids are offered. The winner is the buyer who placed the last bid. The **Dutch auction** is the reverse of the English auction where the auctioneer calls out a very high bid to begin the process and then lowers the bid price until one buyer accepts the price. Dutch auctions tend to transpire faster than other auction formats and are thus used to sell perishable products such as plants and food. **First price auctions** consist of collecting sealed bids from participants and awarding the buyer (or supplier) submitting the highest (lowest) bid the good (government contract). In **second price auctions** buyers also submit sealed bids. But, while the highest bid wins the auction, the winner buyer only pays the second highest bid. eBay has recently begun using a computer mechanism that mimics the second price auction.

See also *multi unit auctions*

Bibliography

Cramton, P. and Ausubel, L. (2002). Demand reduction and inefficiency in multi–unit auctions. Working paper, University of Maryland.

Hirschey, M. and Pappas, J. L. (1995). *Fundamentals of Managerial Economics*, 5th edn. New York: Dryden Press.

McAfee, R. P. (2002). *Competitive Solutions: The Strategist's Toolkit*. Princeton, NJ: Princeton University Press.

McAfee, R. P. and McMillan, J. (1987). Auctions and bidding. *Journal of Economic Literature*, **25**, 699 738.

autocorrelation

Robert E. McAuliffe

When estimating a LINEAR REGRESSION with TIME SERIES DATA the error terms (the residuals in regression analysis) are assumed to be random. A random error is one that, on average, is not related to any preceding errors. When autocorrelation exists, the error term in one period is related to the error terms in previous periods. Such a relation will bias the estimated standard errors of the coefficients in the regression because the EXPECTED VALUE of the error term this period and the error term last period will not be zero. That is,

$$E(\varepsilon_t, \varepsilon_{t\ 1}) \neq 0$$

where ε_t is the error term for period t, and $\varepsilon_{t\ 1}$ is the error term one period ago.

Autocorrelation can be positive, in which case the error term this period is likely to be above average (zero) if the error term last period was above average (zero). When autocorrelation is negative, the error term this period is likely to be below average (zero) if the error term last period was above average (zero). First order autocorrelation means that the error term this period is related, on average, to the error term one period ago. Evidence of first order autocorrelation is provided by the DURBIN–WATSON STATISTIC. An error term with first order autocorrelation is represented as

$$\varepsilon_t = \rho \varepsilon_{t\ 1} + u_t$$

where ε_t is the error term for period t, ρ is the autocorrelation coefficient (which may be positive or negative but must be less than one in absolute value), and u_t is a random error. Second order autocorrelation occurs when the error this period is related to the error last period and the period before that. Higher orders of autocorrelation are also possible.

When the residuals from the regression are autocorrelated, it means that there are persistent errors in explaining the dependent variable with the fitted regression equation. As a result, autocorrelation may indicate that the regression is misspecified and that a significant explanatory variable is missing from the regression equation. Procedures such as the Cochrane–Orcutt method can correct autocorrelation in the regression equation but this should be regarded as a second best solution if additional explanatory variables have not been considered.

See also *time series forecasting models*

Bibliography

Greene, W. H. (2002). *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.

Gujarati, D. N. (2002). *Basic Econometrics*, 4th edn. New York: McGraw-Hill Irwin.

Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.

Wooldridge, J. M. (2002). *Introductory Econometrics: A Modern Approach*, 2nd edn. Cincinnati, OH: South-Western Publishing.

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Industrial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

average total cost

Robert E. McAuliffe

This measures the total economic costs of production per unit produced in the short run and is also referred to as short run average cost. Average total costs include the opportunity cost of capital employed (that is, the normal, risk adjusted rate of return on capital), so a firm operating on its average total cost curve is earning zero ECONOMIC PROFIT (*see* OPPORTUNITY COSTS). Since economic profits are the signal for ENTRY and EXIT from the industry, the average total cost curve represents a bench mark curve in the short run for predicting whether entry or exit will occur (*see* LONG RUN COST CURVES; SHORT RUN COST CURVES).

Average total costs in the short run consist of average fixed costs and average variable costs. FIXED COSTS are those costs of production which do not vary with the level of output and are fixed in the SHORT RUN. Therefore average fixed costs (fixed costs per unit produced) decrease as the quantity produced increases. Variable costs are those costs which vary with the level of output produced such as labor, material inputs, etc. Average variable costs are the variable costs per unit produced and will decrease initially but will eventually increase because of DIMINISHING RETURNS to the variable inputs. Thus the "typical" average total cost curve is U shaped, representing decreasing per unit costs initially as more units are produced, but reaching a minimum and then rising as output rises per period.

See also *average variable costs; economic profit; fixed costs; short run cost curves*

average variable costs

Robert E. McAuliffe

Variable costs are those costs which vary with the level of output produced by the firm in the SHORT RUN and average variable costs are total variable costs per unit produced. If the firm did not produce output in a given period, these costs would not be incurred. The variable costs of production will be affected by the per unit cost of each variable input in production (for example, hourly wage rates for workers), the productivity of the inputs in production, and the production technology available to the firm.

The average variable cost curve is important for short run decisions when the price received for producing output is so low that the firm may choose not to produce. A firm should shut down its operations when, in the short run, it cannot earn revenues sufficient to pay its average variable costs. The firm has no choice regarding its FIXED COSTS in the short run since these costs must be paid whether or not the firm shuts down. Therefore, fixed costs should have *no* effect on the firm's short run decisions. However, the firm does not have to pay the variable costs of production in the short run, so if operating the plant costs more than the revenues earned, the firm should shut down and simply pay its fixed costs. For a perfectly competitive firm, the shutdown point occurs where MARGINAL COST equals average variable cost (that is, when average variable cost is at a minimum). At any price below this point, revenues earned from operations will fail to cover the costs of operations. The firm will have greater losses if it operates and should shut down.

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Industrial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Martin, S. (2001). *Advanced Industrial Economics*, 2nd edn. Oxford: Blackwell.

B

backward integration

Robert E. McAuliffe

A firm which buys one of its suppliers or chooses to produce inputs for itself has integrated backwards in the production process. Such a decision may be motivated by concerns that supplies might be interrupted or because of ASSET SPECIFICITY problems that prevent successful contractual negotiations with existing independent suppliers. Firms may also choose to integrate backwards for strategic benefits such as improving product quality or lowering costs. Backward integration gives the acquiring firm more control over its input supplies but it also requires more careful attention from management and coordination with the upstream firm. The upstream firm frequently wants to obtain input supplies at a lower cost, but if the market for these inputs is competitive, the firm's profits will not be maximized by arbitrarily lowering the price of inputs (see Keat and Young, 2002; Porter, 1980; Shughart, Chappell, and Cottle, 1994).

See also *opportunity costs; vertical integration*

Bibliography

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.
Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.
Shughart, W. F., Chappell, W. F., and Cottle, R. L. (1994). *Modern Managerial Economics*. Cincinnati, OH: South-Western Publishing.

bankruptcy

John Edmunds and Roberto Bonifaz

A firm or individual which is unable to meet its financial obligations passes into a status called bankruptcy after a court of law issues an order declaring it bankrupt. Bankruptcy may be voluntary, if the firm or individual petitions the court to be placed into bankruptcy status. Or it may be involuntary, if creditors of the firm or individual petition the court to declare the entity bankrupt and place it under the stewardship of a trustee. An entity might seek voluntary bankruptcy to gain time to work out a plan to meet its obligations. Creditors might seek to have an entity declared bankrupt to protect its remaining assets from being squandered, dissipated, or stolen; they might also try to force an entity into bankruptcy because the court and the trustee will pay the claims on the bankrupt in order of priority, not according to the whims or preferences of the entity.

Firms may go bankrupt because of illiquidity, or because of insolvency. A firm suffers from illiquidity if, for example, its only asset is a parcel of land worth $100, and its only liability is a debt of $20 payable immediately. The firm is solvent because its asset is worth more than its liability; but it is bankrupt because it does not have the cash to pay the debt that is due. That firm would probably seek bankruptcy voluntarily, to keep the debtholder from getting the parcel of land, to the detriment of the stockholders. Another firm would be liquid but insolvent if its only asset were $50 in cash and it had debts of $100. That firm might be pushed into bankruptcy by its creditors, so that the creditors would share the $50 in order of priority, and

not in the order that the management might choose if it were allowed to continue managing the firm.

The bankruptcy courts have some latitude in settling claims against the bankrupt enterprise. The current practice is to promulgate a settle ment that gives every class of claimant some thing, with only relative priority to the senior claims. Prior to 1978 in the US, the "absolute priority" rule was in effect; that rule gave the senior creditors 100 percent of their principal and interest before giving anything to junior creditors. In recent years the bankruptcy courts have made settlements that give some payment or future claims to every class of creditors, in cluding stockholders.

Bibliography

Brigham, E. (1995). *Fundamentals of Financial Manage ment*, 7th edn. New York: Dryden Press.

barriers to entry

Robert E. McAuliffe

Barriers to entry describe the disadvantages of potential entrants relative to established firms in an industry. Barriers to entry play an important role in determining the structure of an industry, such as the number of firms and the size distri bution of firms (*see* MARKET STRUCTURE). There is some controversy regarding the proper definition and consequences of barriers to entry, however. For the purposes of evaluating the likelihood of entry into any specific industry and the basic structure of an industry, these controversies are less significant. As will be clear below, the differences are more important in the areas of social welfare and public policy.

Sources of Barriers to Entry

As initially described by Joe Bain (1956), barriers to entry allow established firms to raise prices and earn long run ECONOMIC PROFIT without causing new firms to enter the industry. In per fectly competitive markets, short run economic profits attract entry which in the long run causes prices and profits to fall to normal, risk adjusted, competitive levels (*see* ENTRY; PERFECT COM PETITION). If there are barriers to entry in an industry, economic profits will persist even in the long run. For Bain and his followers, higher long run economic profits indicate higher bar riers to entry, all else equal. Bain considered ECONOMIES OF SCALE, PRODUCT DIFFER ENTIATION, absolute cost advantages of estab lished firms, and capital requirements as sources of barriers to entry.

Established firms in an industry have absolute cost advantages when they can obtain resources at a lower cost than potential entrants. This advantage can occur when established firms have exclusive access to important inputs, or have unique assets (such as a prime location, manufacturing process, etc.) that allow them to produce at a lower cost than potential entrants. Under these conditions, established firms would be able to charge prices above their marginal costs and earn economic profits without at tracting entry. If accounting methods valued these superior assets at their true market value (or opportunity cost), the advantages could disappear (*see* OPPORTUNITY COSTS). For example, a firm which enjoys lower costs because of a unique location could sell that loca tion for a higher price. In this case, the firm's location is an asset which is undervalued relative to the true market value of that asset. If the firm properly accounted for the higher value of the asset, the firm's return on assets would correspondingly fall and its costs would be higher.

Bain considered economies of scale as a bar rier to entry because any potential entrant would have to build a large plant in order to compete with established firms. Bain argued large scale entry was inherently more risky and difficult to finance and thus reduced the likelihood of entry. Any firms that entered at a scale below the MINIMUM EFFICIENT SCALE (MES) would have higher costs relative to established firms, and if entrants chose a plant the size of the MES or larger, they would likely provoke retaliation by the established firms. Economies of scale may also create barriers to entry because they require substantial capital investment. Absolute capital requirements refer to the amount of capital ne cessary to successfully enter an industry and launch a product. If capital markets are imper fect, entrants may have difficulty obtaining

sufficient credit or may have to pay higher inter est rates because of the greater risk.

Product differentiation may also confer ad vantages to established firms because entrants would have to compete in marketing the product in addition to producing it. Entrants may also need to overcome consumer loyalty to estab lished brands and this would increase the costs and risks of entry. If an entrant must advertise more to make consumers aware of the product, these higher costs increase the risk of entry be cause they cannot be recovered if the attempt to enter the industry fails. Entrants may also find it difficult to obtain access to distribution chan nels when competing against established firms offering similar products.

Schmalensee (1982) and Porter (1980) also identified SWITCHING COSTS as a potential barrier to entry. These are the costs incurred by consumers when switching from an estab lished, well known brand to a new brand. These costs would appear to be particularly im portant when new products, such as software, require training. Even so, the existence of these costs creates incentives for software companies to ease the transition with special menus and help files for users of competing products. Thus an entrant may be able to reduce these costs to its customers.

Another barrier to entry suggested by Dixit (1979) is actually a barrier to exit: SUNK COSTS. If a firm cannot recover investments when it leaves an industry, then the firm incurs sunk costs and has less incentive to enter that indus try. Sunk costs are greater the more specific those assets are to the industry or firm because there are fewer alternative uses for specific assets (*see* ASSET SPECIFICITY). Advertising outlays to introduce a product are sunk if an entering firm ultimately fails.

CONCEPTUAL QUESTIONS REGARDING
BARRIERS TO ENTRY

George Stigler (1968) disagreed with the notion that economies of scale were a barrier to entry when both the entrant and the established firm have the same costs of production. If the market is not large enough to accommodate two or more firms, we could just as easily argue that insuffi cient demand is the barrier to entry and not economies of scale. Stigler then defined entry barriers as those costs incurred by any new en trants into an industry that were not incurred by established firms. For Stigler, economies of scale do not represent a barrier to entry when entrants have access to the same technology and costs.

More recent research by von Weisäcker (1980) and Demsetz (1982) has questioned the basic concept of barriers to entry by focusing on the welfare implications of more entry into an industry. Von Weisäcker suggested that Stigler's definition of barriers to entry should apply only when the additional costs borne by an entrant imply a misallocation of resources from a socially efficient optimum. For example, when there are many firms and economies of scale, society would be better off with *less* entry because add itional firms incur more fixed costs and prevent the full exploitation of scale economies. But Demsetz questions the usefulness of the distinc tion between industry insiders and outsiders al together. After all, PROPERTY RIGHTS are legal barriers to entry but economists do not consider them to be uniformly anticompetitive. If con sumers prefer established products because it is costly for them to experiment with new brands or acquire more information, is that necessarily a barrier to entry?

In a world of IMPERFECT INFORMATION and TRANSACTIONS COSTS, Demsetz suggests that society may want to encourage investment in trademarks and ADVERTISING. An estab lished firm with a good credit history will be able to borrow more cheaply than a new entrant with no record. Although Stigler might argue that the higher cost of credit to an entrant could be a barrier to entry, does society want to dis courage good credit histories? Or consider the case of a product produced by a patented pro cess. The patent grants a legal MONOPOLY to the producer of this product which will likely result in lower output and a higher price than the MARGINAL COST of production. Society would benefit from increased production of this prod uct, but changing the patent laws would not necessarily be beneficial to society since there would now be a reduction in the incentives to invent. In fact, Demsetz argues that there would now be a barrier to entry in the market for inventing new products. According to Demsetz, changing property rights will change the mix and value of resources in society in ways that

may not increase social welfare. In Demsetz's view, public policy to eliminate or reduce barriers to entry will have unanticipated effects on property rights that may impose greater costs than benefits.

Gilbert (1989) more recently focused on the advantages enjoyed by established firms. He prefers the term mobility barriers to entry barriers, reflecting the fact that resource allocation will be more or less efficient as capital and other resources are more or less mobile in moving from one industry to another. A mobility barrier exists when an established firm in an industry has economic RENTS (positive or negative) as a result of being in the industry. If a firm earns higher profits in an industry *after accounting for the opportunity cost of the resources employed* than it could earn in another industry (*see* OPPORTUNITY COSTS), then that firm has an incumbency rent. This definition considers the opportunity costs of scarce factors of production since an incumbent firm could leave the industry by selling its assets. Absolute cost advantages described above would not necessarily create a mobility barrier since established firms must consider the opportunity cost of those resources responsible for their advantage. However, as Gilbert notes, if the value of the asset is specific to the owner, such as investments in human capital, then the market value (opportunity cost) may not reflect the true value to the established firm and the firm could earn an incumbency rent.

Geroski, Gilbert, and Jacquemin (1990) argue that an advantage of using incumbency rents to define barriers to entry is that these rents can be measured. Furthermore, the height of the barrier to entry is measured by the size of the incumbency rent. Unfortunately, it is difficult in practice to determine the opportunity cost of factors of production and this makes incumbency rents difficult to implement. Geroski et al. also note that STRATEGIC BEHAVIOR could create a barrier to entry since established firms can take actions which make entry less appealing to an entrant. For example, a threat to reduce prices if a new firm enters a market may also deter entry (*see* LIMIT PRICING). They also note that the existence of factors that may impede entry does not imply that these factors reduce efficiency or require policy action.

Fisher and McGowan (1983) questioned the use of accounting profits as a measure of monopoly power in industry studies (*see* ACCOUNTING PROFIT). They asserted the proper measure of economic profitability is the INTERNAL RATE OF RETURN for any investment in an asset. Since firms invest in a number of assets, the internal rate of return will be a weighted average of these internal rates of return, all of which are forward looking. It is very unlikely that the accounting rate of return will equal the appropriate economic rate of return, and in their simulations Fisher and McGowan showed that the relationship between accounting profits and economic profits was very poor.

For managers attempting to evaluate the conditions of entry into an industry, any of the sources above may represent a barrier to entry (see Porter, 1980) and the controversies in the literature are less important. But for economic policy analysis, such as antitrust policy, or for social welfare analysis, the debate in the literature has an important effect on how markets should be evaluated (*see* ANTITRUST POLICY (US)).

Bibliography

Bain, J. S. (1956). *Barriers to New Competition*. Cambridge, MA: Harvard University Press.

Demsetz, H. (1982). Barriers to entry. *American Economic Review*, **72**, 47 57.

Dixit, A. (1979). A model of duopoly suggesting a theory of entry barriers. *Bell Journal of Economics*, **10**, 20 32.

Fisher, F. M. and McGowan, J. J. (1983). On the misuse of accounting rates of return to infer monopoly profits. *American Economic Review*, **73**, 82 97.

Geroski, P., Gilbert, R. J., and Jacquemin, A. (1990). *Barriers to Entry and Strategic Competition*. New York: Harwood.

Gilbert, R. J. (1989). Mobility barriers and the value of incumbency. In R. Schmalensee and R. D. Willig (eds.), *Handbook of Industrial Organization*. New York: North-Holland.

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.

Schmalensee, R. (1982). Product differentiation advantages of pioneering brands. *American Economic Review*, **72**, 349 65.

Stigler, G. J. (1968). *The Organization of Industry*. Chicago: University of Chicago Press.

von Weisäcker, C. C. (1980). A welfare analysis of barriers to entry. *Bell Journal of Economics*, **11**, 399 420.

basic market structures

Robert E. McAuliffe

The basic product market structures in econom
ics are PERFECT COMPETITION, MONO
POLISTIC COMPETITION, OLIGOPOLY, and
MONOPOLY. Perfect competition represents
the benchmark for welfare analysis where there
are numerous buyers and sellers, none of whom
is large enough to affect the market price if they
leave the market. Any firm that wishes to enter
the industry or exit may do so at little cost. The
product is homogeneous (no PRODUCT DIF
FERENTIATION) and consumers and producers
are completely informed. These conditions dis
cipline producers and consumers so that neither
has any influence over price. The firm facing a
perfectly competitive market sells as much as it
can at the market price and cannot compete for
consumers other than on price. If the firm were
to charge a higher price, consumers would im
mediately switch to other suppliers whose iden
tical products are perfect substitutes. Thus the
demand curve facing a perfectly competitive
firm is perfectly (or infinitely) elastic and per
fectly competitive firms, unable to affect the
market price, are said to be *price takers* (*see*
ELASTICITY). In addition, since each firm is
so small relative to the market and has so many
rivals, there is no advantage from efforts to
anticipate competitors' reactions or engage in
STRATEGIC BEHAVIOR.

When products are differentiated, informa
tion is imperfect, entry or exit is costly, or the
number of sellers or buyers is small, markets are
imperfectly competitive. Firms in these markets
generally have some control over price and there
fore face downward sloping demand curves.
Prices in these markets may exceed the MAR
GINAL COST of production and the market may
not be efficient in allocating output. Producers
may also fail to be technically efficient since they
face less pressure to keep costs as low as possible
(*see* IMPERFECT INFORMATION; TECHNICAL
EFFICIENCY; X EFFICIENCY). Under imper
fect competition firms may compete for custom
ers on other dimensions than simply price.
Firms may compete through product differenti
ation, ADVERTISING, strategic behavior, and
other means.

When firms compete in monopolistic compe
tition, each produces a unique, differentiated
product. Entry into the market is free, and
firms advertise and pursue research and devel
opment to further differentiate their products.
Free entry may drive profits down to zero, but
recent research in spatial models indicates that
long run economic profits are possible if existing
firms can choose their locations and deter
entry (Eaton and Lipsey, 1978; *see* ECONOMIC
PROFIT). EXCESS CAPACITY may exist in
monopolistically competitive markets since
firms will produce less output than required for
minimum average cost. These markets may gen
erate too much or too little product differenti
ation depending on the strengths of two
opposing effects. When a new firm enters the
market with a new brand, it cannot acquire all of
the consumer surplus from the new product.
Since it is socially optimal to introduce a new
product if the social benefits outweigh the social
costs, there may be too little product differenti
ation when each firm cannot appropriate all of
the CONSUMER SURPLUS generated by its
brand. The opposing effect is that any new
product is likely to steal consumers away from
existing firms. Since these consumers were al
ready served and stealing them represents a
negative externality to the other firms that the
entrant does not consider, there may be a ten
dency for too much product differentiation. The
net effect depends on which of these two forces
is stronger (Dixit and Stiglitz, 1977; *see* EXTER
NALITIES).

In markets where producers are oligopolists,
each firm reacts to its rivals' strategies and so
strategic behavior becomes important. Profit
maximizing firms must consider how their com
petitors will respond when determining their
best strategy (*see* GAME THEORY). Since each
firm's expectations about the reaction of rivals
can be modeled in a variety of ways, predictions
about the behavior and performance of oligopol
ists will depend upon the model. If the oligopol
ists were to collude on prices, a cartel would exist
and the firms could collectively act as a monop
oly (*see* CARTELS). If there is no COLLUSION
between the firms, a noncooperative oligopoly
exists. In models of noncooperative oligo
poly, there may be COURNOT COMPETITION

where each firm believes its rivals will keep their production levels (quantities) constant whatever its choice of output, or firms could compete through research and development expend itures, advertising, prices, product differenti ation, or other means with varying expectations about competitors' responses in each case (see Lambin, 1976).

When there is a single producer of a product, that firm has a monopoly in the market. The monopolist maximizes profits by restricting output and raising prices until marginal revenue equals marginal cost. This normally enables the firm to earn economic profit, and to maintain its monopoly position, there must be significant BARRIERS TO ENTRY to prevent other firms from entering the market. These entry barriers may be granted by the government through li censing or patents, they may exist because of ECONOMIES OF SCALE (that is, the firm is a NATURAL MONOPOLY), or they may be created by the firm itself through strategic behavior. Even a natural monopoly may not be sustainable if it can be profitable for entry at some price and cost combinations (see Sharkey, 1982).

Bibliography

Dixit, A. and Stiglitz, J. (1977). Monopolistic competition and optimum product diversity. *American Economic Review*, **67**, 297 308.

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Eaton, B. C. and Lipsey, R. G. (1978). Freedom of entry and the existence of pure profit. *Economic Journal*, **88**, 455 69.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Lambin, J. J. (1976). *Advertising, Competition, and Market Conduct in Oligopoly Over Time*. Amsterdam: North-Holland.

Sharkey, W. W. (1982). *The Theory of Natural Monopoly*. Cambridge: Cambridge University Press.

beta coefficient

Robert E. McAuliffe

In portfolio theory, a company's beta coefficient measures the variability of that company's returns relative to the variability of the returns in the market and it helps determine the com pany's cost of capital within the CAPITAL ASSET PRICING MODEL (or CAPM). An investor who can always receive the risk free rate of return, r_f, will want to be compensated for additional risk from holding any stock. This is the risk premium for that stock and represents the additional return required by investors over the risk free rate. If investors have diversified their port folios, the additional risk incurred by investing in stock j is not the variance of that company's returns (since diversification will reduce some of the risk), but rather how that company's returns relate to the rest of the portfolio. The beta coef ficient measures the additional risk a given stock adds to a portfolio and for stock j it is defined as:

$$B_j = (\text{cov}(r_j - r_f, r_m - r_f))/\text{var}(r_m - r_f)$$

where cov() is the COVARIANCE of the risk premium of stock j with the risk premium of the market portfolio, var() is the VARIANCE of the risk premium of the market, r_f is the risk free rate of return, r_j is the return to stock j, and r_m is the return to the market portfolio. The beta coefficient is simply the estimated coefficient from a LINEAR REGRESSION of the risk pre mium for stock j against the risk premium for the market as a whole.

A company can determine its cost of equity capital by determining the risk premium diversi fied investors will require to add stock j to their portfolios. This is given by:

$$\text{cost of equity capital} = r_f + \beta_j(r_m - r_f)$$

where β_j is the beta coefficient for company j. Since an average stock will move with the market, an average stock should have a beta coefficient of 1. This means that the cost of equity capital for such a company equals the market return: if the market rises or falls by 1 percent, the company's stock will also rise or fall by 1 percent. If a company's beta coefficient is 2.5, the stock will rise or fall two and one half times more than any rise or fall in the market. This would be a more risky stock to add to the portfolio and a portfolio comprised of such stocks would be considered aggressive.

Bibliography

Berndt, E. R. (1991). *The Practice of Econometrics*. Reading, MA: Addison-Wesley.

Brigham, E. F. and Gapenski, L. C. (1997). *Intermediate Financial Management*, 5th edn. New York: International Thomson Publishing.

Shughart, W. F., Chappell, W. F., and Cottle, R. L. (1994). *Modern Managerial Economics*. Cincinnati, OH: South-Western Publishing.

brand name

Robert E. McAuliffe

An established brand name makes a product easy to identify and can reduce consumers' search costs for their best choice. Firms producing a product may invest in establishing a brand name to assure consumers of the quality of that product. This can be particularly important when there is ASYMMETRIC INFORMATION in the market so that consumers cannot easily determine product quality on inspection (as with EXPERIENCE GOODS and CREDENCE GOODS).

Firms can invest in a brand through ADVERTISING, product improvements, and sales efforts. Brand names represent durable investments that show the producer's COMMITMENT to the product over the long term. If consumers were unable to determine product quality and firms could not indicate product quality through SIGNALING, then competition on price could create a LEMONS MARKET where only low quality producers remained. Brand names provide information to consumers which helps ensure that quality levels are maintained even in markets when there is IMPERFECT INFORMATION. Consumers are more willing to pay a price premium for brand names the more sensitive they are to quality differences between products and the greater the costs of acquiring information.

Establishing or maintaining a brand name may also be part of a competitive strategy focused on quality leadership in the market (see Douglas, 1992; Porter, 1980). In such a case the firm tries to achieve competitive advantage and continually earning higher profits by becoming the leading quality producer in the market for the product. The firm's sales efforts, product design, and marketing efforts must all be focused on achieving high quality in the eyes of consumers.

Bibliography

Akerlof, G. A. (1970). The market for lemons: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, **84**, 488 500.

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Klein, B. and Leffler, K. (1981). The role of market forces in assuring contractual performance. *Journal of Political Economy*, **89**, 615 41.

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.

Shughart, W. F., Chappell, W. F., and Cottle, R. L. (1994). *Modern Managerial Economics*. Cincinnati, OH: South-Western Publishing.

budget constraint

Robert E. McAuliffe

When choosing the products which will give the consumer the most satisfaction (*see* UTILITY MAXIMIZATION), a consumer's expenditures cannot exceed his or her budget. In practice this constraint may include the resources available through borrowing and selling assets, but in most textbook examples the consumer is limited to this period's income. For example, to illustrate a consumer's decision to buy two products, X and Y, where the consumer's entire income will be exhausted on these two goods, the budget constraint would be:

$$I = P_X^* X + P_Y^* Y$$

where I is the consumer's income, P_X is the price of good X, X is the quantity of X consumed. If the consumer spent all of her income on good X, then Y would be zero and the maximum amount of X which could be consumed is $X = (I/P_X)$. The budget constraint shows the combinations of the products the consumer could feasibly buy given market prices and the consumer's income

and shows the rate at which a consumer is *able* to substitute purchases of X for Y. Utility maximization requires that the rate at which consumers are *able* to substitute products just equals the rate that they *desire* to do so given their tastes. The consumer's desired rate of substitution is his or her MARGINAL RATE OF SUBSTITUTION and utility is maximized when the consumer's indifference curve is just tangent (equal) to the budget constraint (*see* INDIFFERENCE CURVES).

When the budget constraint is plotted with good Y on the vertical axis and good X on the horizontal axis, the vertical intercept of the budget constraint will be (I/P_Y) (where the consumer spends all of her income on good Y), and the slope of the budget line will be $-P_X/P_Y$. An increase in P_Y will make the budget constraint flatter, reducing the height of the vertical intercept (since less of Y can be purchased), while an increase in P_X will make the budget line steeper and will reduce the horizontal intercept.

Bibliography

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Publishing.

bundling

Robert E. McAuliffe

When firms offer two or more products together at a price below what would be charged for each product separately, the firm is practicing product bundling. Firms will benefit from bundling commodities when they can gain more of the CONSUMER SURPLUS from their customers. In this respect, bundling is a form of PRICE DISCRIMINATION where consumers will sort themselves (or self select) on the basis of how much they prefer a product. For example, quantity discounts are a form of bundling where additional units cost less when purchased together. Those consumers who value the product

the most (that is, have the highest RESERVATION PRICE) will be willing to buy the larger quantity at a discount, although their total expenditures on the good will be higher than if they purchased the smaller size.

The firm profits by raising the prices of the individual products and then choosing the appropriate bundled price to win those consumers who desire both products but are unwilling to pay for both separately. Consumers then self select by purchasing the individual products or the bundled product according to their total valuation of the good.

Bibliography

Adams, W. J. and Yellen, J. L. (1976). Commodity bundling and the burden of monopoly. *Quarterly Journal of Economics*, **90**, 475 98.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Martin, S. (2001). *Advanced Industrial Economics*, 2nd edn. Oxford: Blackwell.

business entities

S. Alan Schlacht

Various legal entities exist for operating a business enterprise. Each of these has advantages and disadvantages from a legal perspective.

The simplest method of operating a business is the **sole proprietorship**. Any person who engages in business without forming any other entity is a sole proprietor. Since a sole proprietorship is nothing more than the person involved, that person has unlimited liability for all business and personal debts. It does not matter that he may use a tradename to do business, he remains liable. In addition, raising capital depends upon the financial ability of the individual. No state approval is required to start or operate a sole proprietorship, no sharing of profits occurs, and taxation is relatively easy.

A **general partnership** is an association of two or more persons who carry on as co owners a business for profit. The word "persons" means that any legal entity may be a partner, including corporations. The partnership may be through an express agreement, either oral or written, or it

may arise through the acts of the parties showing that they intend a partnership to exist. A sharing of profits is the primary evidence that a partner ship is intended. The partnership agreement normally governs the rights and duties of the partners. In the absence of a contract, state law will control. Virtually all states have adopted the Uniform Partnership Act in an attempt to stand ardize partnership law.

Once formed, a partnership is a legal entity for certain limited purposes. It may enter contracts, sue and be sued, and own property. However, it is not a legal entity for tax purposes since the partners pay tax on the income earned. A part nership is easy to form and operate, can raise capital from each partner, and pays no taxes. The major disadvantage of a partnership is that each partner is jointly and severally liable for the debts of the partnership. Thus, each partner is responsible for acts of the other partners per formed within the scope of the partnership. Third parties may collect their entire debt from any partner and leave the partner to seek collec tion from the other partners. In addition, when ever a partner joins or leaves the partnership, whether voluntarily or not, the partnership is dissolved and a new one results. However, the problems of dissolution may be minimized by having a comprehensive agreement.

A **limited partnership** is a variation of the general partnership. The difference is that in addition to the general partner(s) in the firm, limited partners are also present. These partners contribute capital and share in the profits, but they do not participate in the operation of the partnership business. In return, their liability is limited to their contribution. A limited partner ship is formed in a manner similar to a corpor ation. A certificate of limited partnership is obtained from the state and a partnership agree ment is entered. The sale of limited partnership interests may require registration with the Se curities and Exchange Commission (SEC) and each state involved. A limited partnership is operated in the same manner as a general part nership except that the limited partners do not participate for fear of losing their limited liability.

A **corporation** is a legal entity with many of the same legal rights as any person. It can sue and be sued, own property, and enter contracts.

It even has many of the same constitutional rights as people, including equal protection, due process, and freedom of speech. A corpor ation is formed by filing articles of incorporation with the state. A certificate of incorporation is issued showing that the state recognizes the cor poration as a person. However, the corporation must then follow all of the laws pertaining to corporations in order to maintain its identity as a separate legal entity apart from its stock holders and employees. This allows the stock holders, who are the owners of the company, to avoid liability for acts of the corporation.

Corporations are owned by stockholders who contribute capital. They in turn elect directors to serve on the board. The board of directors is responsible for setting major corporate policy and hiring the officers. The officers are the cor porate agents responsible for the daily operation of the corporation and its business. As long as the corporation maintains a separate identity, it alone is responsible for its debts and liabilities. Although a corporation is more difficult to form and operate, this limited liability is a major factor when choosing a business entity. A disadvantage is the double taxation of corporations. Since a corporation is a person, it must pay tax on its income. The corporation may then distribute profits to its stockholders in the form of divi dends, which are then taxed as income to those individuals. One way to avoid this tax burden is to seek Internal Revenue Service (IRS) approval for the corporation to file as a Subchapter S corporation. This designation allows the com pany to be treated as a partnership for tax pur poses. However, the corporation must meet several requirements, including only one class of stock and no more than 35 stockholders. In addition, many corporations must register their stock with the SEC since stock is a security.

Many states have enacted corporate codes that allow for different types of corporations. No state allows professionals, such as physicians or attorneys, to avoid professional liability by forming a corporation. Thus, professionals must form **professional corporations** (noted after their names as PC) instead of business corporations. Professional corporations are op erated similarly to regular corporations except that the stockholders, who must all be within the same profession, remain liable for their

malpractice. In an effort to reduce the administrative burden on small companies, states have also created the **close corporation**. Since all of the stock is "closely" held by a few stockholders, the close corporation statutes treat these companies more like a general partnership but without the unlimited liability. The IRS continues to treat them as corporations for tax purposes.

Many states have also attempted to attract business by allowing **limited liability companies** (LLC) and **limited liability partnerships** (LLP). An LLC is very similar to a general partnership except that the owners (who are called members) do not have liability for debts of the LLC. If an LLC is treated by its members as a partnership rather than a corporation, then the IRS will tax it as a partnership and allow it to avoid double taxation. An LLP is designed for professionals who want to limit liability. A few states have crafted their LLP laws to allow professionals to limit liability in certain circumstances. The trend appears to be toward a limitation of liability for professionals in this area.

Other entities exist, although they are utilized less frequently. A **joint venture** is when two or more business entities come together for a single or limited business venture. The parties are treated legally as partners for that venture only. A **cooperative** is an association (which can be incorporated) that is formed to provide a service to its members, such as a farm cooperative. A **syndicate** is when several persons pool their money, usually within a second business entity, in order to finance a business venture. A business **trust**, or sometimes a real estate investment trust, occurs when investors turn over their money or property to a trust which then manages those resources for profit and distributes profits to the members (beneficiaries).

Finally, many business entities choose to operate a **franchise**. A franchise is *not* a legal entity, but rather it is a method of doing business. The franchisor gives the franchisee information, materials, and support for a fee. The Federal Trade Commission regulates franchising (*see* FEDERAL TRADE COMMISSION ACT).

Bibliography

Moye, J. E. (1998). *The Law of Business Organizations*, 5th edn. Clifton Park, NY: Delmar Learning.

C

capital

Robert E. McAuliffe

Productive assets which yield benefits over time are called capital or economic capital. Capital can be *tangible* (or physical), such as buildings, plant, and equipment, or *intangible*, such as the goodwill of a firm or human capital. A firm may invest resources in developing a BRAND NAME which will produce benefits to the firm over time and which represents an intangible, goodwill asset to the firm. Training programs a firm might provide to employees will develop their skills and make them more productive. This investment in human capital, if successful, will also yield benefits over time but is not part of the physical assets (or capital) of the firm. It is also important to distinguish capital assets from financial assets. Capital assets are used in the production process while financial assets are paper that may repre sent the capital assets but are not the capital assets themselves. For example, a company which is in BANKRUPTCY has all of its capital assets intact. The problem is that the value of the financial assets has declined relative to fi nancial liabilities.

When firms make decisions to invest in cap ital, the benefits are expected to accrue over a period of time in the future and these are CAPITAL BUDGETING decisions. A firm adds to its capital assets by making decisions to invest in new assets over time. Since a capital invest ment decision is a forward looking decision, the expected benefits from such an investment must be estimated and are subject to considerable UNCERTAINTY. Firms use a variety of methods to determine whether a capital investment is worthwhile, but the basic economic principle is that the marginal (expected) benefits from the project should exceed the marginal costs (*see* MAKE OR BUY DECISIONS; NET PRESENT VALUE CRITERIA).

Bibliography

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

capital asset pricing model

James G. Tompkins

The capital asset pricing model (CAPM) is an equation which specifically links the expected return of a security to its underlying risk. A unique feature of the CAPM is that the risk component is fully reflected in only one param eter known as beta (*see* BETA COEFFICIENT). The CAPM equation is:

$$\bar{r}_i = r_f + \beta_i(\bar{r}_m - r_f)$$

where:

$\bar{r}_i$ = expected return on security i
r_f = return for holding a risk free security
β_i = beta of security i
$\bar{r}_m$ = expected return for holding the market portfolio

Although credit for the CAPM goes to Sharpe (1964), Lintner (1965), and Mossin (1966), the work is considered an extension of the portfolio selection research by Markowitz (1952) and Tobin (1958).

Since CAPM is a model that links expected return and risk, it is important to discuss why beta is the parameter that quantifies priced risk in any security. By definition, risk is about unexpected events. For example, if a security's actual return ($\tilde{r}_i$) is always equal to its expected return ($\bar{r}_i$), then it is intuitive that such a security has no risk (since the expected always occurs), and would therefore provide an investor with the risk free rate of return (r_f). The total risk or VARIANCE (σ_i^2) of a security is therefore calculated as the EXPECTED VALUE of the square of its unexpected returns. Formally, $\sigma_i^2 = E(\tilde{r}_i - \bar{r}_i)^2$, where E denotes mathematical expectation.

If an investor had no choice but to accept the risk of security i, then the variance would capture the relevant risk for pricing purposes. However, Markowitz (1952) had the insight that when a single security is part of a portfolio, there exists the potential for risk cancellation. For example, if you held stock in both an umbrella and ice cream factory, it is easy to imagine that an unexpectedly hot and dry summer would boost ice cream sales while depressing umbrella sales. Hence, the risk of unexpected weather would be less if both stocks were held in a portfolio instead of holding them individually. The concept that risks may cancel is known as *diversification*. Thus, if the total risk or variance inherent in a single security can be reduced simply by costlessly holding it as part of a well diversified portfolio, then the relevant risk for pricing purposes is the risk that the single security contributes to this well diversified portfolio. By definition, the risk of a single security which may be diversified is known as its *unique* or *non systematic* or *diversifiable* risk. The risk component of a single security which cannot be diversified is known as *market* or *systematic* or *non diversifiable* risk. Analogous to the variance of a single security quantifying its total risk, non diversifiable risk is quantified by the *beta* of the security.

In the set of risky securities, Markowitz (1952) illustrated a frontier of portfolios such that each of them had the greatest return for the given level of portfolio risk. The upper portion of the graph represents the *efficient frontier*. Assuming that an investor prefers higher expected returns and less risk, all investors will choose a portfolio somewhere on the efficient frontier. (Other technical assumptions include that each asset is infinitely divisible, and that all investors have common time horizons and common beliefs about the investment opportunity set and their expected returns.)

If we include a riskless security in the above analysis, by definition it will have a portfolio standard deviation of zero. Assuming investors can borrow and buy the riskless asset in unlimited quantities, all investors will choose a convex combination of the riskless asset and the risky portfolio denoted as S in figure 1. This line is known as the *capital market line* and illustrates the investor's *separation principle*. This principle states that investors are able to separate two specific decisions. The first is to calculate the set of efficient assets represented by the efficient frontier as well as the point of tangency between the riskless asset and the efficient frontier (point S). The second decision is to determine which combination of the portfolio S and the riskless asset an investor will choose. If an investor has a low degree of risk tolerance, she will invest some of her funds in the riskless asset and some in portfolio S. If she has a high degree of risk tolerance, she will borrow at the risk free rate in addition to using her funds to invest in portfolio S.

In the set of risky assets, if everyone holds portfolio S, then this must also be the market portfolio. To appreciate why beta reflects the risk that a single security contributes to a well diversified portfolio, it is necessary to understand the calculation of portfolio risk. This risk is calculated as a weighted average of the vari
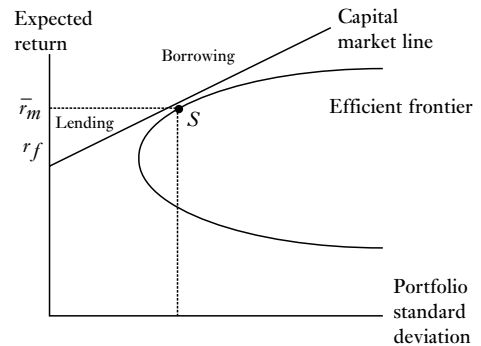

Figure 1   Optimal portfolio selection

ance of each security in the portfolio plus all the COVARIANCE terms between the securities. Therefore, if w_1 and w_2 are the fractions of your wealth held in stocks 1 and 2, respectively, and the covariance between the two stocks is denoted by σ_{12} then portfolio STANDARD DEVIATION (σ_p) is calculated as the square root of portfolio variance:

$$\sigma_p = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1 w_2 \sigma_{12}}$$

Notice that with two securities, there are two variance and two covariance terms. Similarly, with three securities, there are three variance and six covariance terms. In short, when there are n securities in a portfolio, there are n variance and $(n^2 - n)$ covariance terms. If we have an infinite number of securities, by definition we hold the market. Hence as n approaches infinity, it can be shown that the covariance of a single security with the market dominates the risk that this security contributes to a well diversified portfolio. The covariance that a single security i has with the market standardized by the variance of the market is known as its *beta*. Hence $\beta_i = \frac{\sigma_{im}}{\sigma_m^2}$ where σ_{im} is the covariance between security i and the market. Intuitively, if the beta of a stock is 2, this means that when the market rises by 1 percent and the risk free rate does not change, then the stock will rise by 2 percent.

Conceptually, beta is key to the CAPM equation since it quantifies the non diversifiable risk inherent in any single security. Since it is this risk which cannot be costlessly diversified away, and since we assume that investors expect a higher rate of return for higher risk stocks, it must be this non diversifiable risk that is priced. We are now in a position to understand the intuition behind the CAPM equation.

Whereas the capital market line is derived from the set of risky portfolios and the riskless asset, the security market line shown in figure 2 represents the trade off between risk and return for a single security.

The security market line illustrates several important points. The first is that the beta of the risk free rate and the market (M) must be 0 and 1, respectively. The second is that this line must be upward sloping since investors require higher returns for higher risk. The third point is
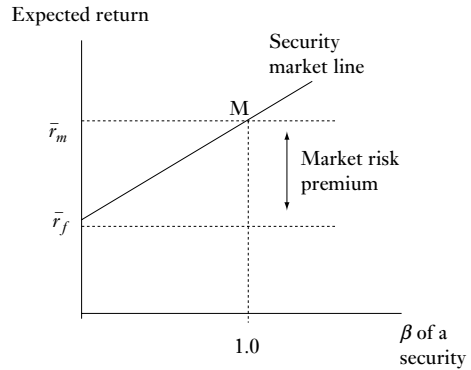
Expected return

that in EQUILIBRIUM, this must be a straight line. If this were not true, it would mean that the price of risk (ratio of expected return premiums to risk) differed across securities, which would provide the motivation to simultaneously buy risk at a low price and sell it at a high price. Hence, if the risk premium of any security is the difference between its expected return and the risk free rate of return, then it must mean, for example, that

$$\frac{\bar{r}_A - r_f}{\beta_A} = \frac{\bar{r}_B - r_f}{\beta_B} = \frac{\bar{r}_c - r_f}{\beta_C}$$

The security market line graphically illustrates the CAPM. With the risk free rate as the intercept, the market risk premium ($\bar{r}_m - r_f$) as the slope, and the expected return and beta of a security as the dependent and independent variables, respectively, we obtain the CAPM equation:

$$\bar{r}_i = r_f + \beta_i(\bar{r}_m - r_f)$$

It is important to point out that the CAPM is not without its critics. For example, Roll (1977) has argued that the model is impossible to test since it is a prediction about expected returns, while an empirical test would have to use actual returns. In addition, there are alternative theories of asset pricing such as arbitrage pricing theory pioneered by Ross (1976). Nevertheless, the CAPM is a model which is widely used today by practitioners throughout the world. For example, a component in the weighted average COST OF CAPITAL calculation includes the cost

of equity capital which can be determined using CAPM. Since the cost of capital is key to making major capital budgeting decisions, the CAPM implicitly plays an important role. Another example is that numerous investment reports including *Value Line* show a stock's beta to assess its priced risk. In short, the CAPM is an excellent example of theory born in the academic world that has successfully integrated into many practical uses in the business environment.

Bibliography

Lintner, J. (1965). Security prices, risk and maximal gains from diversification. *Journal of Finance*, **20**, 587 615.

Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance*, **7**, 77 91.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica*, **34**, 261 76.

Roll, R. (1977). A critique of the asset pricing theory's tests. Part I: On the past and potential testability of the theory. *Journal of Financial Economics*, **4**, 129 76.

Ross, S. A. (1976). The arbitrage theory of asset pricing. *Journal of Economic Theory*, **13**, 341 60.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, **19**, 425 42.

Tobin, J. (1958). Liquidity preference as behavior towards risk. *Review of Economic Studies*, **25**, 65 86.

capital budgeting

Vickie L. Bajtelsmit

The process of making investment decisions involving fixed assets is called capital budgeting. Once a particular long term investment has been identified, the firm's management must estimate the expected cash flows from the project and the timing of those cash flows over the life of the project (*see* CASH FLOW). The relevant cash flows are the investment outlays that will be required at the outset of the project, including expected increases in net working capital, and the annual net incremental cash inflows, that is, the difference between the firm's net cash flows with and without taking on the project. Thus, for example, the analysis would not include SUNK COSTS, but would include OPPORTUN

ITY COSTS and EXTERNALITIES associated with the investment.

The net operating cash flows generally consist of sales revenue minus expenses and taxes plus any tax savings due to the allowable DEPRECIATION deduction. At the end of the project, there may also be cash flow associated with the salvage value of the used fixed asset. Since it is difficult to make accurate long term forecasts of sales, expenses, and other associated cash flows, the potential for error in the forecasts may be great, particularly for large, complex investments. Therefore, the next step in the capital budgeting process is to assess the riskiness of the projected cash flows. The management will then determine the appropriate COST OF CAPITAL to be used in determining the NET PRESENT VALUE (NPV) of the project. Alternatively, the management may make their decisions based on INTERNAL RATE OF RETURN (IRR), in which case risk will be incorporated into the analysis by adjusting the "hurdle rate," or minimum acceptable IRR. When the projected cash flows are riskier, the management will increase the hurdle rate to account for the additional risk. The investment opportunity schedule is a graphical depiction of the firm's opportunities in terms of IRR and dollars of new capital raised. Since stockholder value will be maximized by choosing the projects with the greatest return, this schedule will be downward sloping.

When a firm is faced with many different project alternatives, the optimal capital budget is determined by choosing the set of projects that maximizes the net worth of the firm. As the firm increases its total level of new investment, it must raise new capital to support this investment. The intersection of the increasing marginal cost of capital curve and the decreasing investment opportunity schedule will determine the optimal capital budget, that is, the level of investment and the specific set of projects that will maximize the firm's value.

Although NPV and IRR are the most commonly used decision methods in capital budgeting, there are several alternative techniques that are sometimes used in practice. For example, some managers use payback analysis, in which projects are evaluated based on the time required to recoup initial cash expenditures. This method is generally inferior to NPV and

IRR since it ignores cash flows that occur after the break even point. Another alternative decision tool is the "profitability index" (PI), a ratio of the present value of the cash inflows to the present value of the cash outflows. A PI that is greater than one is equivalent to NPV greater than zero and the criteria are applied similarly.

Bibliography

Brigham, E. F. and Gapenski, L. C. (1997). *Intermediate Financial Management*, 5th edn. New York: International Thomson Publishing.

Gitman, L. J. (1995). *Foundations of Managerial Finance*, 4th edn. New York: HarperCollins.

capital markets

Don Sabbarese

Capital market securities are transferred through a mix of markets and financial intermediaries. The efficient flow of capital market funds is critical for financing real asset growth and economic expansion. Financial securities with maturities greater than one year are traded in capital markets and include debt obligations as well as common and preferred equity. The market for new debt and equity issues is called the primary market. Outstanding debt and equity securities are traded on the secondary market.

Debt securities offer investors interest payments and/or a par or face value payment at maturity. Debtholders can also sell their securities on the secondary market prior to its maturity. Common equity offers investors an expected dividend stream, the potential for capital gains, and voting rights. Preferred stock includes the provision that their dividends must be paid before those to common stockholders.

Various debt and equity securities offer investors a range of features designed to meet individual preferences for risk and return. These securities may differ with respect to default risk, interest rate risk, liquidity, and taxability. Callable features, convertible features, sinking funds, adjustable rates, tax exemptions, and derivatives are designed to meet changing investor concerns. The capital market infrastructure of intermediaries provides suppliers and demanders of funds with critical decision making information and facilities necessary for allocating funds based on risk return preference.

Commercial rating companies provide credit analysis and default risk ratings on corporate debt securities. Moody's Investment Services and Standard and Poor's are the most widely used bond rating systems. Large institutional investors and investment banks also operate their own credit analysis system.

Investment banks are intermediaries in the primary market for new issues. They collect, assemble, interpret, and disseminate information. New issues prices are set and sometimes guaranteed by investment banks. Brokerage firms and dealers are intermediaries in the secondary market, where outstanding securities are traded. Brokers, through their membership in a central exchange such as the New York Stock Exchange, buy and sell listed stocks for their clientele, to whom they also act as advisers. Unlisted stocks and debt obligations are traded in over the counter markets such as the National Association of Securities Dealers (NASDAQ). Government securities dealers are specialists who both buy and sell government debt of all maturities.

The capital markets have experienced major changes in the last 30 years. First, innovations in computer technology have led to more sophisticated trading strategies. Other innovations such as pass throughs and other asset backed securities evolved along with other securities and related derivatives such as collateralized mortgage obligations and stripped mortgaged backed securities to create new lower risk securities with a broader appeal to investors. Additional concerns over risk management have encouraged the growth of futures and options markets. Finally, small investors' preference for lower risk redirected funds through mutual funds and pensions. As a result, trading systems today are designed to meet the needs of institutional investors rather than those of small investors.

To small investors, institutional investors offer financial services and products, risk reduction, and liquidity as well as reduced costs of contracting. Commercial banks, savings and loans, insurance companies, finance companies, pension funds, and mutual funds intermediate billions of small investor dollars into capital markets based on risk return preferences and

the nature of their business. Commercial banks invest in a wide range of US government secur ities, federal agency notes, and municipal se curities so as to meet regulatory restrictions for safety and liquidity. In contrast, savings and loans specialize in mortgages. Life insurance companies select their capital market holdings so that their maturities will match the company's expected policy payouts. Likewise, pension funds tailor their investments to debt securities and equity which match predictable pension payouts. Mutual funds invest in a full range of securities with a special mix of securities in each fund.

Bibliography

Fabozzi, J. F. and Modigliani, F. (2002). *Capital Markets, Institutions and Instruments*, 3rd edn. Englewood Cliffs, NJ: Prentice-Hall.

capitalism

Gilbert Becker

Every society must select an economic system for its use. The primary importance of such a system is that it allocates society's scarce re sources. Here, three fundamental questions are resolved. The first concerns *what* goods and services should be produced in the upcoming period with the resources that are available. For example, should more encyclopedias and fewer cars be produced, or should society spend more resources this year on building more homes? This question recognizes the fundamental prob lem (for business and society) of trade offs which occur when scarcity exists (*see* OPPOR TUNITY COSTS).

The second question is *how* to produce these goods. Most goods can be produced with differ ent production methods, using varying combin ations of labor, CAPITAL, and other inputs. Here it is in the firm's best interest, for PROFIT MAXIMIZATION, and also society's best inter est, for UTILITY MAXIMIZATION, that the resources are used efficiently (*see* EFFICIENCY). Finally, an economic system must decide *for whom* these goods should be produced (i.e., who will be the ultimate recipients of these goods).

Capitalism is one of several economic systems (e.g., socialism, feudalism) which resolve these questions. Capitalism is based on several funda mental principles, including:

1 *Private ownership of property*, where individ uals maintain legal claim over the goods that they purchase in the economy, and more importantly, claim ownership rights to the productive resources such as land, capital, and their own labor. (This is in contrast to socialism, where widespread government or collective ownership of industries is in effect.)
2 *Free enterprise*, wherein consumers have the right to purchase those goods and services that they choose, and resource owners have a similar right to choose which goods and ser vices to produce. Moreover, managers are given wide latitude in decisions ranging from production processes to store location and hours of operation.
3 A third feature is a basic *reliance on markets*. Here the forces of supply and demand (*see* LAW OF DEMAND) are used to generate prices for goods and inputs. These prices are central to the decision making of con sumers and managers alike, and thus are critical in resolving the three basic questions cited above. For example, growing market demand for a good may allow for a higher price for that product, inducing entrepre neurs to move more of their resources into the production of that good and away from another. Similarly, the relative prices of the different resources needed to produce a good (*see* ISOQUANT ISOCOST CURVES), and which consumers are willing and able to pay for that good (*see* DEMAND FUNC TION), determine *how* the good is to be produced, and *who* will ultimately own it.
4 COMPETITION is a final element that is cen tral to capitalism. Rivalry involving price, ADVERTISING, product attributes (*see* PRO DUCT ATTRIBUTES MODEL; PRODUCT DIFFERENTIATION; PRODUCT VARIETY), and other elements of STRATEGIC BEHAV IOR is essential to achieving the important social goals of allocative and productive effi ciency, and equity (*see* MARKET PERFORM ANCE; PERFECT COMPETITION).

Although widely practiced in countries around the world, capitalism is not universally preferred in every country. Nor is unfettered capitalism fully embraced by all who live in countries (e.g., the US) where it is the prevailing system. Among those who prefer capitalism, differing economic ideologies exist. The type of, and extent to which, government intervention is advocated distinguish these ideologies in large part. See Carson, Thomas, and Hecht (2002), who offer a comparison of the competing ideological views as they pertain to several current issues.

The eighteenth century Scottish philosopher and economist Adam Smith is often considered as one of the champions of capitalism, and represents the more **conservative** economic ideology. Smith's seminal work, the *Wealth of Nations* (1776), helped to develop basic economic principles such as the division of labor and ECONOMIES OF SCALE, but is more importantly a call for free enterprise and open markets, especially with respect to international trade. He also saw individual ownership of property as a natural right, and competition being the driving force behind innovation, higher standards of living, and greater social wellbeing in general. This last idea was described by Smith as a situation wherein managers, in an effort to use their resources to produce goods with the most value (and thereby make the most profit for the firm), were simultaneously acting as if guided by "an invisible hand" toward promoting the best interests of society as a whole. These principles are the foundation of today's conservative economic philosophy, perhaps best represented by the works of Friedman (1962, 1983), which call for maximum economic freedom and exclusive use of markets to allocate resources.

More **liberal** economic thinking (in the US) also fundamentally believes in the system of capitalism, but within certain restraints imposed by government. For example, while a market system is preferred and free enterprise largely applauded, the liberal economic thinker is more open to situations in which government determines resource allocation, through programs ranging from Medicare health insurance for the elderly to rent control on housing property. Moreover, while liberals do not generally endorse widespread government ownership of business, they are more willing to accept situations such as the postal service, where the government owns considerable resources used in delivering the mail. Finally, and perhaps most importantly to managers, liberal economic philosophers, while believing in the benefits of competition, often fear that insufficient competition exists in some markets (*see* OLIGOPOLY). They therefore advocate government regulation of certain business practices, such as ADVERTISING, and promote the enforcement of stricter competition policy laws and guidelines (*see*, for example, FEDERAL TRADE COMMISSION ACT; HORIZONTAL MERGER GUIDELINES). One important twentieth century liberal economic thinker, J. K. Galbraith (1958, 1973), argued that consumers lack sovereignty and are manipulated by large corporations, and that free enterprise and a *laissez faire* approach to markets tend to promote the interests of the powerful and big business, rather than those of society as a whole.

Interestingly, liberal economic thinkers often cite one familiar passage of Adam Smith's work (1776: 128) which reads, "people of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public or some contrivance to raise prices." This statement, perhaps an important insight into oligopolistic interdependence and COLLUSION, demonstrates for liberal economic thinkers the need for antitrust laws. Nonetheless, their more conservative counterparts cite several passages in Smith's work indicating the futility and even obstructive nature of much government intervention. As a result of these differing philosophies, most capitalist systems are mixed, wherein elements of pure capitalism are combined with varying amounts of government oversight of the use of productive resources and the sale of goods.

Managers often see government intervention into markets as undesirable and unnecessary interference with free enterprise and the market system. Yet it is valuable for managers to understand that not all government action lies in conflict with capitalism or business interests. In particular it should be recognized that government is needed to, and does, support capitalism, by acting as judge and referee, especially where market failure may otherwise occur.

One such area exists where the ownership rights to property are in question. Here the role of government in determining, protecting, and enforcing property rights is essential to the smooth functioning of markets. For example, in the case involving Napster Inc., the courts established that the ownership rights of produ cers of music had been violated by the free dissemination of copywritten material over the Internet. Similarly, another area where there exists legitimate public concern, in the eyes of liberal and conservative thinkers alike, occurs where EXTERNALITIES exist. Here, third parties, not directly involved in the purchase or sale of a good, may benefit or be harmed none theless. A prominent example is that of pollu tion, wherein air or water pollution resulting from the production or consumption of a good in one region of a country may influence the utility of the public in another region. Here, the lack of private property rights (e.g., to air) causes the market failure (too much polluted air is produced). Government action may be neces sary to correct the problem, by assigning prop erty rights or imposing penalties or restrictions on polluters. Coase (1960) challenges this gov ernment involvement, arguing that the parties may negotiate a corrective agreement by them selves, if TRANSACTIONS COSTS are small (*see* COASE THEOREM).

See also *antitrust policy (US); Clayton Act; structure conduct performance paradigm*

Bibliography

Carson, R., Thomas, W., and Hecht, J. (2002). *Microeco nomic Issues Today: Alternative Approaches*, 7th edn. Armonk, NY: M. E. Sharpe.

Coase, R. (1960). The problem with social costs. *Journal of Law and Economics*, 3, 1 44.

Colander, D. (2004). *Microeconomics*, 5th edn. Boston: McGraw-Hill Irwin.

Friedman, M. (1962). *Capitalism and Freedom*. Chicago: University of Chicago Press.

Friedman, M. (1983). *Bright Promises, Dismal Perform ance: An Economist's Protest*. Sun Lakes: Thomas Horton.

Galbraith, J. K. (1958). *The Affluent Society*. Boston: Houghton-Mifflin.

Galbraith, J. K. (1973). *Economics and the Public Purpose*. Boston: Houghton-Mifflin.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. New York: Modern Library.

cartels

Robert E. McAuliffe

When firms make formal agreements to set prices or output levels they may form a cartel to reduce competition between them. Cartels, such as OPEC in oil production and DeBeers in dia monds, are illegal in the US under its antitrust policy (*see* ANTITRUST POLICY (US)), which prohibits any agreements between firms which have the effect of reducing COMPETITION. Cartel members are cooperating oligopolists who may decide to maximize joint profits, divide the market, or set other goals depending on the number of members and a variety of economic forces. Under US antitrust laws, any agreement to fix prices is illegal *per se* and no defense of the agreement is allowed. The European Union (EU) prohibits COLLUSION between firms when the sole purpose is to reduce competition, but ex emptions may be granted if the agreement satis fies specific public goals (see Martin, 1994).

Salop (1986) identified three requirements for firms to successfully collude. First, the firms must agree on the goals the cartel is to achieve; second, the firms must achieve those goals; and third, the firms must maintain the cartel agree ment as market conditions change over time. Since firms face different circumstances in terms of their demand and cost conditions, col lusive agreements may not be easily achieved. For example, agreements to maximize joint profits would require that output be produced by the lowest cost producer. This could mean that some members of the cartel produce very little output. The cartel would then need to make side payments to these firms, but they might not be satisfied with the payments they receive or with their status in the cartel. Furthermore, once the cartel is established, all firms have incentives to cheat on the agreement by selling additional output below the cartel price since that price will exceed the MARGINAL COST of production. The incentive to cheat makes it more difficult to maintain the cartel agreement (Stigler, 1968). Successful cartels must be able to detect and

punish cheating members, yet punishing chea ting firms by increasing output and forcing the price down will punish all members if the prod uct is homogeneous. The incentives to cheat are greater when firms' marginal costs of production do not increase much with increases in output, when the percentage of total costs that are FIXED COSTS is high, and when the probability of de tection is small. For a given price, the first two conditions mean greater profits for each add itional unit produced while the last concerns the likelihood of punishment. Agreements are easier to reach and enforce in industries with a few, large sellers; that is, industries which are highly concentrated (Stigler, 1968; *see* MARKET STRUCTURE).

Cartel stability is also threatened by potential ENTRY into the market, since the higher prices charged by established firms create incentives for new entrants, so viable cartels must be pro tected by BARRIERS TO ENTRY. The greater the differences between producers in products, costs, or goals, and the more unpredictable are market conditions, the more difficult it will be to agree, achieve, and maintain collusive arrange ments.

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Indus trial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Martin, S. (1994). *Industrial Economics*, 2nd edn. New York: Macmillan.

Martin, S. (2001). *Advanced Industrial Economics*, 2nd edn. Oxford; Blackwell.

Salop, S. C. (1986). Factors that (credibly) facilitate oli gopoly co-ordination. In J. E. Stiglitz and G. F. Math ewson (eds.), *New Developments in the Analysis of Market Structure*. Cambridge, MA: MIT Press, pp. 265 90.

Stigler, G. J. (1968). *The Organization of Industry*. Chi cago: University of Chicago Press.

cash flow

Robert E. McAuliffe

A firm's cash flow represents the difference be tween the firm's cash outlays and receipts over a given period of time. Estimates of cash flows for any investment project under consideration are crucial for CAPITAL BUDGETING decisions based on either the NET PRESENT VALUE or INTERNAL RATE OF RETURN criteria. Since future cash flows for an investment are not known with certainty, managers should calculate expected cash flows based on the probability they will occur. In addition, any estimated net present value or internal rate of return should be tested to determine how sensitive the results are to slight changes in the underlying assumptions (such as changes in the probabilities of the cash flows or in the discount or hurdle rate). All cash flows should be estimated on an after tax basis.

Estimated cash flows should also be adjusted for anticipated INFLATION, particularly since a company's revenues may not adjust to inflation as its costs. For example, a firm in a very com petitive market may not be able to raise prices at the same rate as inflation, yet the firm's costs may rise with the inflation rate (due to cost of living adjustments in workers' contracts). In such cases managers may want to separately adjust the revenue and expenditure sides of the estimated cash flows for inflation.

Bibliography

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Reider, R. and Heyler, P. B. (2002). *Managing Cash Flow: An Operational Focus*. New York: John Wiley.

certainty equivalent

Robert E. McAuliffe

As the term suggests, the certainty equivalent of a risky decision is the amount an investor would require to make her indifferent between accepting that sum (with certainty) or accepting the chance of a risky decision. If an investor is risk averse, then the certainty equivalent will always be less than the expected payoff of a risky decision. The greater a person's RISK AVERSION, the lower the certainty equivalent the person will require for a given risky payoff. For example, an investor who is just indifferent between paying $100 for a stock that could pay

$300 or nothing, with 50 percent probability for each, has a certainty equivalent of $100.

The certainty equivalent factor is the ratio of the certainty equivalent sum to the expected risky payoff. In the example above, the investor was indifferent between a certain $100 and an expected payoff of $150, so the certainty equivalent factor is 66.67 percent. The cost of one dollar in risky returns is 66 cents of certain returns for this investor. An investor who knows this factor can then adjust the expected future returns from a given investment by this factor to adjust the investment for risk when making portfolio allocation decisions, for example. Riskier investments should have lower certainty equivalents and therefore will have lower certainty equivalent factors.

Bibliography

Hirschey, M. (2003). *Fundamentals of Managerial Eco nomics*, 7th edn. Cincinnati, OH: South-Western Publishing.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

ceteris paribus

Robert E. McAuliffe

In economic analysis, many factors can affect a given variable. As an analytical device, econo mists employ the ceteris paribus assumption (Latin for "all else equal") to isolate the effect of one of the independent variables on the de pendent variable. For example, the demand for a product will be affected by the price of that product, the prices of substitute products (*see* SUBSTITUTES), COMPLEMENTS, and income among other possible variables. To find the effect of a change in the product's own price on the quantity demanded, the assumption "ceteris paribus" is used to derive a theoretical conclu sion. This makes the statement about the rela tion between the quantity demanded and price a *conditional* statement: *if* no other factors affect ing demand are changed, *then* a rise in the prod uct's price should cause the quantity demanded to fall. Clearly, if the conditions are changed (for example, incomes increase and this is a normal

good), then we may not expect to see a decrease in the quantity demanded even if the price is higher because all else was not equal (constant); *see* NORMAL GOODS.

Bibliography

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

Clayton Act

Gilbert Becker

The US Clayton Act was written in 1914, in part to clarify the legislatures' intentions concerning phrases such as "in restraint of trade," which were included in the 1890 SHERMAN ACT. The Clayton Act specifies a number of business practices which are proscribed. The Act includes sections on PRICE DISCRIMINATION, TYING, and mergers, among others. Each type of con duct is identified as a violation of the law, but is not treated as *per se* illegal as the courts had ruled in some Sherman Act cases involving behavior such as price fixing. Instead, the Clayton Act considers these competitive strategies as being illegal only if "the effect may be to substantially lessen competition or tend to create a monop oly"; see Breit and Elzinga (1996) and Jorde, Lemley, and Mnookin (1996) for a more com plete statement of this Statute and for details concerning cases cited below. Thus a so called *rule of reason* approach is used, whereby the adverse consequences (and potential benefits) of the business practices are examined.

Section 2 of the Clayton Act, and its 1936 revision named the Robinson–Patman Act, con cerns price discrimination, which is the act of selling the same product to different customers or customer groups for different prices. The law was originally designed to diminish the use of geographic PREDATORY PRICING. Here a firm having MARKET POWER in one GEOGRAPHIC MARKET would set an artificially low price (below its cost) for its products in another geo graphic market in an attempt to drive out its rivals and extend its market power to the new

area. This could be done by temporarily covering the firm's losses with the economic profits earned in the initial market (*see* ECONOMIC PROFIT). The revised law also focuses on protecting smaller buyers in a market from some of the advantages of large (e.g., chain store) rivals in that market.

Managers should be aware of several legally accepted defenses against charges of price discrimination which have been established. First, price differences are allowed where they exist due to a justifiable difference in costs. Thus differences may be allowed due to service and transportation costs or quality differences in the product or quantity discounts resulting from cost savings. Second, price differences resulting from a good faith effort to meet a competitor's price are allowed. Third, price discrimination may be defended on the basis of the product in question having lost its marketability due to obsolescence or due to its perishable nature. Finally, the lack of a "substantial lessening of competition" or injury to competitors serves as a possible defense. As a result, price discrimination that exists in the airline, movie, and magazine industries, for example, is not seen as a violation of the Act.

Section 3 of the Clayton Act concerns illegal tying arrangements and exclusive dealing contracts. Tying involves a situation in which the seller offers a product (the tying good) to a customer, conditional on the customer's additional purchase of a second (tied) good. The principal concern of the law is the extension of market power from the tying good market into the market for the tied good. For example, the A. B. Dick Company, manufacturer of copier machines, ink, brushes, and other accessories, originally tied the purchase of these accessories to that of the machine. The company held patents and therefore some market power in the market for copiers, but the courts saw the market for the accessories as being essentially competitive. The courts feared that the tie allowed the company to charge prices above the competitive level for the accessories and thereby inappropriately extend its legally acquired market power in the copier market into the market for accessories.

Here again managers should be aware of the circumstances where tying contracts are allowed. Once again, where competition may not be "sub-

stantially lessened" the tie may be allowed. As a result, one defense involves the demonstration of either:

1   a lack of market power in the market for the tying good, or
2   a limited impact on the market for the tied good.

Second, a tie is allowable where obvious production ECONOMIES OF SCALE are in evidence.

A third defense occurs when the protection of the tying good's quality or the goodwill of the firm necessitates a tie. This precedent was established in the 1960 case of *US* v. *Jerrold Electronics Corporation*; see Stelzer (1986). Here Jerrold, a supplier of community television antenna systems, was allowed by the courts to require the purchasers of the system to also purchase their maintenance contracts, as the systems were new and sophisticated and required specially trained service personnel. In an earlier case, though, in which the International Salt Co. tied the purchase of its own rock salt to the sale of its salt dissolving machines, the courts ruled that the tie was illegal, in part because the company's machines were not, as Justice Jackson put it, "allergic to salt of equal quality" produced by another firm; see Breit and Elzinga (1996). A fourth (absolute) defense arises by demonstrating that the tie between two goods is voluntary or optional to the customer rather than being an absolute requirement. Finally, the recent antitrust lawsuit involving the Microsoft Corp. has presented two additional defenses. The first involves the existence of NETWORK EXTERNALITIES in the industry. The second arises by questioning whether or not the two goods in question (i.e., Microsoft's Windows operating system and its Internet Explorer web browser) are truly separate goods (*see* MICROSOFT ANTITRUST CASE).

The practice of exclusive dealing is related to tying in that a customer is "tied" to a particular supplier. Deals between manufacturers and materials suppliers, for example, may connect the two parties via long term contracts where the former agrees to be exclusively supplied by the latter. Similarly, contracts between manufacturers and retailers may require the latter to purchase a product or set of products exclusively

from one manufacturer. The key antitrust concern in both cases here is market foreclosure, in which there is an adverse effect on competition since some rivals are excluded from open competition in the market. In the first case, an exclusive contract with a large manufacturer may foreclose other suppliers from competing for sales to an important buyer in the market. In the second case, one example of foreclosure involves the practice of "full line forcing," in which a manufacturer of one line of products, for example oil and gasoline, requires retailers to purchase other products such as tires and other accessories exclusively from this seller as well. This contract may foreclose other accessories suppliers from an important retail outlet. The courts have held that an exclusive dealing contract is illegal if it has a substantial anticompetitive effect. This has typically been measured by examining the dollar volume of sales affected by the contract. As a result, such deals may be illegal even when less than 10 percent of the market may be foreclosed.

Antitrust enforcement agencies and the courts have also recognized that important benefits may arise for both buyers and suppliers involved in exclusive contracts. A manufacturer may gain from having a guaranteed source of steady supply of materials, which enables better planning, lower backup inventories, and a smoother flow of output (and perhaps lower prices for its customers) as a result. The supplier gains from a steady source of demand, which minimizes promotional expenditures, diminishes the uncertainty of long term capital expenditure planning, and also establishes a steady flow of revenues. As a result, the standards applied in the rule of reason approach to resolving these cases have varied over time and across cases.

Section 7 of the Clayton Act focuses on mergers. Here again mergers are outlawed only when their effect "may be to substantially lessen competition or tend to create a monopoly." Perhaps no other area of antitrust policy receives as much scrutiny and debate over the proper interpretation of the law. Three times in the last quarter century guidelines concerning merger policy have been written and substantially rewritten by the federal government in an effort to clarify the existing administration's views on the proper interpretation of Section 7 and the

circumstances under which the appropriate federal agencies will bring antitrust policy action against merging parties (*see* ANTITRUST POLICY (US); HORIZONTAL MERGER GUIDE LINES; MERGER GUIDELINES, 1992–7; MONOPOLY).

There exist three types of mergers for policy makers to consider. These are:

1 horizontal mergers, involving direct rivals who are producing a similar product in the same geographic market;
2 vertical mergers, between firms involved at different stages of raw material supply, production, or distribution of the same good (*see* VERTICAL INTEGRATION); and
3 conglomerate mergers, which deal with firms competing in essentially separate product markets.

Of greatest concern to policy makers are horizontal mergers, as this type of merger reduces the number of competitors, increases the market share of the acquiring firm in the merger, and thereby potentially changes the MARKET STRUCTURE and increases the firm's market power. Vertical mergers are also of concern, most often when a merging firm with market power in one stage would now be able to restrict competition in another stage. For example, in the AOL–Time Warner merger, concerns were voiced that Time Warner, a leading owner of cable distribution systems, could deny access to other Internet service providers, thereby giving AOL increased dominance in its market. Much less concern has been given to conglomerate mergers unless the acquiring firm, in the absence of a merger, may have entered on its own, thereby increasing the number of competitors.

Active antitrust policy concerning mergers was pursued during the 1960s and 1970s during which time the federal government instituted strict guidelines indicating that an antitrust challenge will ordinarily occur against the horizontal merger of firms having relatively small market shares. At the same time, the Supreme Court supported this position by initiating the *incipiency doctrine*, wherein it indicated that a trend (resulting from mergers) toward higher industry concentration would be stopped in its earliest stages. More recently, especially during the

1980s and 1990s, the extent of antitrust activity against mergers has diminished.

See also *antitrust policy (US); Federal Trade Commission Act; tying*

Bibliography

Blair, R. and Kasserman, D. (1985). *Antitrust Economics*. Homewood: Irwin.

Breit, W. and Elzinga, K. (1996). *The Antitrust Casebook: Milestones in Economic Regulation*, 3rd edn. Fort Worth, TX: Harcourt Brace.

Church J. and Ware, R. (2000). *Industrial Organization: A Strategic Approach*. Boston: McGraw-Hill.

Jorde, T., Lemley, M., and Mnookin, R. (1996). *Gilbert Law Summaries: Antitrust*, 9th edn. Chicago: Harcourt Brace Legal and Professional Publications.

Stelzer, I. (1986). *Selected Antitrust Cases: Landmark Decisions*, 7th edn. Homewood: Irwin.

Waldman, D. and Jensen, E. (2001). *Industrial Organization: Theory and Practice*, 2nd edn. Boston: Addison-Wesley.

Coase theorem

Steven G. Medema

The Coase theorem was developed from Ronald Coase's pathbreaking work in "The problem of social cost" (1960). The theorem has been stated in a number of different forms, but its central thrust is that, if rights are well defined and TRANSACTIONS COSTS are zero, the allocation of resources will be efficient regardless of the initial assignment of rights (*see* PROPERTY RIGHTS). Thus, if a power plant emits pollution which harms the operation of a nearby laundry, it does not matter, from an efficiency perspective, whether the court assigns the power plant the freedom to pollute or the laundry the right to be free from pollution. The ability of the parties to negotiate costlessly will result in the right coming to rest where it is most highly valued.

The Coase theorem is one of the most controversial propositions in modern economics, and understandably so, given its profound implications. For the theorem implies that the structure of law is without effect on the allocation (albeit not the distribution) of resources. Efficiency does not require making the nuisance generator, the tortfeasor, or the contract breacher liable for the harm that they cause. This proposition goes against the grain of a half century of welfare economics, which has steadfastly argued that the efficient resolution of EXTERNALITIES requires that the government place some type of restraint on those who generate the externalities; against this, the Coase theorem asserts that markets can indeed work to resolve externality problems. Because of this, the theorem has become a weapon in the hands of those who would defend the market against the incursions of government. Rather than tax or regulate polluters, government can simply assign rights and let the market work. Furthermore, there is no reason to hold liable those who generate the harm; efficiency will result in either case.

The linchpin of the theorem is the assumption of zero transactions costs, a concept which has been construed both narrowly and broadly. The theorem tells us that private negotiations, the pricing mechanism, or the market (as the case may be) will allocate rights and resources efficiently if there are no barriers to their working. This has justly been called a tautology. On the other hand, it is an interesting tautology for several reasons. First, it tells us that the market can work in situations where it was previously thought to fail, provided that the necessary conditions obtain. Second, it suggests that where markets do fail, one cause may be the presence of transactions costs which are sufficiently high to preclude a negotiated solution. This both adds to our understanding of why inefficient externalities persist and suggests that, where removable impediments to negotiation exist, measures can potentially be implemented to facilitate a market solution. Third, the theorem can be extended to encompass situations where transactions costs are positive: rights will be reallocated through negotiation to higher valued uses so long as the expected gains from such negotiations make it worthwhile to incur the costs. This is perhaps the most important application of the theorem, for, as Coase recognized, and as numerous subsequent commentators have pointed out, transactions costs are omnipresent – a fact that violates the central assumption of the theorem itself.

The effect of this last point is that the market can, in certain instances, work to resolve externality situations. The issue is one of setting up

the proper framework to allow it to do so. The original stimulus for Coase's analysis was his argument that the government should consider setting up a market in broadcast frequencies rather than allocating them by administrative fiat (Coase, 1959). While this idea was roundly rejected for some time after it was made, we have, today, auctions of rights to frequencies upon which new portable telephone technologies will operate. And the current experiments with marketable pollution permits are suggestive of the more general applicability of Coase's contention that the pricing mechanism may be preferable to more traditional externality remedies in certain contexts.

The third point, above, makes particularly clear the link between Coase's analysis in "The problem of social cost" and that in "The nature of the firm" (1937). In the latter essay, Coase undertook to find a rationale for the existence of the integrated firm, including both the hierarchical employer–employee relationship (*see* HIERARCHY) and the decision to make or buy (*see* MAKE OR BUY DECISIONS). Coase suggested that, if market transactions were costless, all production activity would take place through a series of market transactions; the firm would not exist. The firm comes into existence because there are costs associated with transacting through the market, costs which are sometimes higher than those associated with internal organization. The firm, he said, will organize additional transactions internally as long as the cost of doing so is less than the cost of carrying out the same transaction through the market. Thus, while the Coase theorem shows that, from an efficiency perspective, the law has no purpose (apart from some initial assignment of rights), Coase's work in "The nature of the firm" suggested that, if market transactions are costless, the firm has no purpose. Of course the import of this comes from the converse: since market transactions are not costless, the institutional structure (the structure of law and the organizational structure of the firm and of firm vis à vis market) becomes an important part of economic analysis, as these factors play a prominent role in affecting the allocation of resources in society.

This, in turn, brings to the fore the importance of the contracting process and the examination of those factors and forces which influence the process itself and the structure of CONTRACTS. The Coase theorem suggests that all contracts will be fully specified against all eventualities. Observed contracts are very different from this, often leaving open a host of potential actions due, at least in part, to the costliness of contracting over these terms (*see* INCOMPLETE CONTRACTS). The result is that, in many instances, the respective rights of the involved parties are not well defined, exposing agents to potentially adverse actions by their counterparts. The effect of this is to induce the expenditure of resources on various types of safeguarding activities or the devising of mechanisms (such as internal organization as against the market, or one type of internal organizational structure as against another) to lessen the potential for, or effect of, these adverse actions. Consider a situation where A is the sole supplier of an input to B and B does not have readily available alternative sources of supply. B is thus exposed to potentially opportunistic behavior by A. The transactions costs associated with the contracting process may make it difficult or impossible to eliminate this potential within the contract, and A's best response may thus be to produce the input itself (vertically integrate).

Thus, while the Coase theorem is a statement of the virtues of markets, its implications come from the relaxation of its zero transactions costs assumption, which highlights the role that transactions costs play in determining the institutional structure of production and that this institutional structure plays in affecting the costs of transacting.

Bibliography

Barzel, Y. and Kochin, L. A. (1992). Ronald Coase on the nature of social cost as a key to the problem of the firm. *Scandinavian Journal of Economics*, **94**, 19 31.

Coase, R. H. (1937). The nature of the firm. *Economica*, n.s., **4**, 386 405.

Coase, R. H. (1959). The Federal Communications Commission. *Journal of Law and Economics*, **2**, 1 40.

Coase, R. H. (1960). The problem of social cost. *Journal of Law and Economics*, **3**, 1 44.

Coase, R. H. (1988). *The Firm, the Market, and the Law*. Chicago: University of Chicago Press.

Coase, R. H. (1992). The institutional structure of pro-
duction. *American Economic Review*, **82**, 713 19.

Medema, S. G. (1994). *Ronald H. Coase*. London:
Macmillan.

Medema, S. G. (ed.) (1995). *The Legacy of Ronald Coase
in Economic Analysis*. Aldershot: Edward Elgar
Publishing.

Cobb Douglas production function

Robert E. McAuliffe

The Cobb–Douglas production function is fre
quently employed in the economic analysis of
production and costs. This production function
relates output produced to the inputs of produc
tion (*see* PRODUCTION FUNCTIONS). If labor
hours, L, and capital, K, are the only inputs
in production, the Cobb–Douglas production
function is:

$$q = AL^{\alpha}K^{\beta}$$

where A is a coefficient that represents the level
of technology and α and β are coefficients indi
cating how output responds to changes in each
of these inputs (additional inputs can be added
to the function above in a similar fashion). A
convenient feature of the Cobb–Douglas pro
duction function above is that ECONOMIES OF
SCALE in production can be determined by
examining the coefficients α and β. When
there are economies of scale in production,
doubling all inputs will cause output to increase
by more than double. In the case above, if we
double the labor and capital inputs, we would
have:

$$q = A(2L)^{\alpha}(2K)^{\beta} = 2^{(\alpha+\beta)}AL^{\alpha}K^{\beta}$$

Doubling the inputs used in production will
cause output to rise by $2^{(\alpha+\beta)}$, so the returns to
scale in production can be determined by sum
ming the coefficients of the Cobb–Douglas pro
duction function. If the sum of $(\alpha + \beta)$ equals
one, then there are CONSTANT RETURNS TO
SCALE.

Estimates of the Cobb–Douglas production
function can be obtained by a LINEAR REGRES
SION of the logarithm of output against the
logarithm of the relevant inputs in production
(see Berndt, 1991; Maddala, 2001).

Bibliography

Berndt, E. R. (1991). *The Practice of Econometrics*. Read-
ing, MA: Addison-Wesley.

Cobb, C. and Douglas, P. H. (1928). A theory of produc-
tion. *American Economic Review*, **18**, 139 65.

Greene, W. H. (2002). *Econometric Analysis*, 5th edn.
Upper Saddle River, NJ: Prentice-Hall.

Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd
edn. New York: John Wiley.

coefficient of variation

Robert E. McAuliffe

When making decisions under risk, the mean
and STANDARD DEVIATION provide informa
tion about the degree of dispersion, but they
are affected by the size and units of measure
used. If a company were considering two pro
jects with substantially different EXPECTED
PRESENT VALUE, the means and standard de
viations are not directly comparable. The coeffi
cient of variation is a measure of relative risk
and is defined as the ratio of the standard
deviation to the expected present value of the
project:

$$\text{coefficient of variation} = \sigma/\text{EPV}$$

where σ is the standard deviation of returns from
the project and EPV is the expected present
value of the project. Since the expected present
value and the standard deviation are measured in
the same units, the units will cancel and the
coefficient of variation will be unit free. It is a
measure of the risk incurred per dollar of
expected return. The inverse of the coefficient
of variation is EPV/σ which measures the risk
adjusted return.

Bibliography

Douglas, E. J. (1992). *Managerial Economics*, 4th edn.
Englewood Cliffs, NJ: Prentice-Hall.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial
Economics: Economic Tools for Today's Decision
Makers*, 4th edn. Upper Saddle River, NJ: Prentice-
Hall.

collusion

Robert E. McAuliffe

When firms in a market have formal or informal agreements on pricing, output, or other com petitive issues, the firms are in collusion. Anti trust policy in both the US and the European Union (EU) prohibits collusion between com peting firms in a market (*see* ANTITRUST POLICY (US); EU COMPETITION POLICY, 2004). Collusive agreements can be tacit ar rangements whereby firms in an industry follow a price leader when it changes price, or they may be the formal agreements of a cartel (*see* CARTELS). Collusion is more likely in industries which are highly concentrated (see Stigler, 1968; Martin, 1994) and when the costs of organizing firms are lower.

Salop (1986) identified a number of factors which he called facilitating practices that have the effect of promoting information exchange among firms or promoting incentives to adhere to the agreement. Practices which promote in formation exchange help reduce uncertainty about the actions of competitors and therefore help maintain collusive agreements such as ad vance notice of price changes. According to Salop, most favored nation (MFN) clauses in sales contracts help create incentives for firms to keep to the agreements because if one firm attempts to decrease its price, it must do so to *all* of its customers. This makes price reductions more costly and increases the likelihood that competitors will learn of the discounts. Meet the competition clauses have a similar effect in providing pricing information to all competitors. Combining MFN and meet the competition clauses in contracts to buyers commits the estab lished firm to retaliate automatically if another firm in the industry (or a potential entrant) cuts prices and this credible strategy can help main tain the collusive agreement (*see* CREDIBLE STRATEGIES).

Bibliography

Martin, S. (1994). *Industrial Economics*, 2nd edn. New York: Macmillan.
Martin, S. (2001). *Advanced Industrial Economics*, 2nd edn. Oxford: Blackwell.
Salop, S. C. (1986). Factors that (credibly) facilitate oli gopoly co-ordination. In J. E. Stiglitz and G. F. Math ewson (eds.), *New Developments in the Analysis of Market Structure*. Cambridge, MA: MIT Press, pp. 265 90.
Stigler, G. J. (1968). *The Organization of Industry*. Chi cago: University of Chicago Press.

commitment

Robert E. McAuliffe

Commitments that a firm makes can be crucial to the success of its STRATEGIC BEHAVIOR. An established firm may threaten to reduce price if ENTRY occurs, but if the firm cannot commit itself to the price reduction, a potential entrant may not believe the threat, particularly if the established firm would maximize profits by avoiding the price cut. But if a firm invests in capacity to produce, achieves lower costs through the LEARNING CURVE, or includes meet the competition clauses in its sales con tracts, the irreversible nature of these actions commits the firm to behave aggressively should entry occur. The commitment serves to "back up" the threat of retaliation by the established firm, and therefore the behavior of the firm's competitors will be affected. If the firm cannot make commitments, its strategic threats may not be taken seriously (Jacquemin, 1987; Porter, 1980).

Capital expenditures in specialized assets or in brand ADVERTISING can have commitment value because these are SUNK COSTS which cannot be recovered. The more specialized the assets of the firm, the greater the sunk costs if the firm fails (*see* ASSET SPECIFICITY). These costs can be a SIGNALING device to competitors that the firm will compete aggressively to defend its market position. Porter (1980) argues that the success or failure of strategic plans, offensive or defensive, depends on the firm's ability to make commitments that are to its advantage. A firm that makes commitments signals its position and likely reaction to competitors, and therefore forces them to adjust their decisions. When firms can make commitments before a rival does, those firms may also have FIRST MOVER ADVANTAGES. Porter lists a number of visible assets which can serve to communicate a firm's commitment: excess cash reserves, EXCESS

CAPACITY, and extensive research facilities, which can all be used to threaten rival firms.

Commitment may also be necessary to over come TRANSACTIONS COSTS in the market. A firm may be unwilling to undertake investments in specific assets for a customer unless it can be certain that the customer will purchase the output. CONTRACTS are a formal way of estab lishing commitment between parties to a trans action when the obligations can be reasonably specified.

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Indus trial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Jacquemin, A. (1987). *The New Industrial Organization: Market Forces and Strategic Behavior*. Cambridge, MA: MIT Press.

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

competition

Robert E. McAuliffe

Competition has different interpretations in eco nomics and business. Businessmen and women regard competition as a process of rivalry where firms try to gain an advantage through pricing decisions, ADVERTISING, research and devel opment, product QUALITY, and other means. In economics, however, when there is PERFECT COMPETITION firms cannot engage in rivalrous behavior because they are price takers in the market. This difference in the meaning of com petition has created some difficulties in the past.

A market is perfectly competitive if the following conditions hold (see Stigler, 1968; Carlton and Perloff, 2000):

1   The product is homogeneous: the firms in the industry produce goods which are iden tical in the eyes of consumers.
2   Information is perfect: consumers and sup pliers know all offers to buy and sell, and producers also know the returns they could earn if their resources were put to alternative uses.
3   Large numbers of buyers and sellers: this condition is necessary to ensure that no single seller or buyer can affect the market price and means that market participants are price takers.
4   Free ENTRY and EXIT: resources must be free to move to where they have the highest value. If there are BARRIERS TO ENTRY in an industry, then new firms cannot take ad vantage of profitable opportunities when they arise.

Since firms in a perfectly competitive indus try are price takers, their only decision is the level of output they produce. Competition under these circumstances is so intense that any firm which tries to charge a price slightly above the market price will have no sales. There fore perfect competition in economics is more a description of a state of the market than a pro cess. Nevertheless, it has very desirable proper ties in terms of economic EFFICIENCY and represents a benchmark against which other market structures may be compared. Since the choices of firms in a perfectly competitive market are so constrained, interactions between firms do not matter and difficulties modeling STRATEGIC BEHAVIOR are avoided.

Critics such as Hayek (1948) and High (1990) have argued that the assumptions of perfect competition remove the most essential features of competition in real markets. The dynamic process of rivalry between firms leads to a variety of observed market behavior such as advertising, cost cutting, quality improvement, and product innovations. These are the weapons of the com petitive struggle between firms that create improved standards of living in market econ omies and better serve consumers. By focusing on perfect competition, these authors argue that economists may miss important features of the market process. While the point is well made, conventional analyses of markets and com petition do include these forms of price and nonprice competition (see, for example, Carlton and Perloff, 2000; Porter, 1980).

When firms compete in markets that are not perfectly competitive, all of the competitive tools described above may be used by firms to

improve their market position. Are other forms of competition as effective in eliminating economic profits as price competition? Stigler (1968) showed that price competition would be more effective in reducing profits if the marginal cost of changing the nonprice variable (such as advertising or product quality) is larger than the marginal cost of reducing price (*see* ECONOMIC PROFIT; MARGINAL COST).

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Industrial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Hayek, F. A. (1948). *Individualism and Economic Order*. Chicago: Chicago University Press.

High, J. (1990). *Maximizing, Action, and Market Adjustment*. Munich: Philosophia Verlag.

Porter, M. (1980). *Competitive Strategy*. New York: Free Press.

Stigler, G. J. (1968). *The Organization of Industry*. Chicago: University of Chicago Press.

complements

Robert E. McAuliffe

Complements are those products which are consumed together with another product, such as tires or gasoline and an automobile. The services of the automobile cannot be rendered without gasoline, so a consumer's willingness to pay for an automobile will be affected by the cost of gasoline. The prices of complementary goods have a negative effect on the quantity demanded for a product because if the price of a complement (gasoline) rises, the total cost of driving an automobile is now higher and this may reduce the demand for automobiles.

The negative effect of the price of complementary goods on the demand for a product would be observed as a negative cross elasticity of demand (*see* CROSS ELASTICITIES). This means that a given percentage increase (decrease) in the price of a complementary good will cause the demand for the related product to fall (rise) by "Y" percent.

See also *substitutes*

Bibliography

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Publishing.

concentration indices

Robert E. McAuliffe

Concentration indices are summary measures describing elements of MARKET STRUCTURE. Market concentration reflects the number and size distribution of firms in a market. As Stigler (1968) has noted, the measurement of size itself depends upon the purposes for which the measure is to be used. If the goal is to explain deviations of prices or profits from their levels under PERFECT COMPETITION, then measuring firm size by its employment is appropriate when examining the labor market, but less so for the product market where sales are a more appropriate measure. Firm size in the capital market may be better measured by its total assets, while in the market for materials, its purchases may better represent the firm's MARKET POWER.

The choice of a concentration measure should also be determined by the purpose for which it is to be used and several measures of concentration have appeared in the literature. The concentration ratio measures the percentage of industry output produced by the largest T firms (usually T is taken as four or eight). If s_i is the market share of firm i, then the T firm concentration ratio would be:

$$CR_T = \sum_{i\,1}^{T} s_i \qquad (1)$$

for $T = 4$ or $T = 8$. The Justice department used the four firm concentration ratio for its 1968 merger guidelines; an industry was considered highly concentrated if the four firm

concentration ratio equaled or exceeded 75 per cent. Unfortunately, because it focuses only on the largest firms in the industry, the concentration ratio does not describe the distribution of firm sizes throughout the industry, nor does it provide information about the sizes of the individual firms that comprise the top four or eight firms. Concentration ratios are typically reported for *domestic* producers only, so they exclude competition from foreign imports and therefore may overstate the degree of concentration in an industry.

Another frequently used measure is the HER FINDAHL HIRSCHMAN INDEX, or HHI, which is calculated using the market shares of all n firms in the industry, measured as:

$$\text{HHI} = \sum_{i=1}^{n} s_i^2 \qquad (2)$$

The HHI requires more information to compute relative to the concentration ratio and it is consistent with theories of OLIGOPOLY and COURNOT COMPETITION (Stigler, 1968; Jacquemin, 1987). The HHI has a maximum value of 10,000 in the case of a single firm and a minimum value near zero in the case of a large number of very small firms in the market. The HHI can also be written to represent the number of firms in the industry and the variance of their market shares. Thus:

$$\text{HHI} = n\sigma^2 + 1/n$$

where σ^2 is the variance of the firms' market shares in the industry (Waterson, 1984; Jacquemin, 1987).

The entropy index, EI, is less often used as a measure of concentration. It is calculated as the sum of all firms' market shares times the logarithm of their market shares:

$$\text{EI} = \sum_{i=1}^{n} s_i \ln(s_i) \qquad (3)$$

The entropy index gives greater weight to smaller firms than the HHI. Another index, the Linda index (LI), is based on concentration ratios but reveals differences in the sizes of the large firms. The T firm Linda index is calculated as:

$$\text{LI} = \frac{1}{T(T-1)} \sum_{i=1}^{T-1} \frac{T-i}{i} \frac{\text{CR}_i}{\text{CR}_T - \text{CR}_i} \qquad (4)$$

When all T firms are the same size, this index has the value of $1/T$ and when the $(T+1)$th firm is substantially smaller than the preceding T firms, the value of LI_{T+1} will be much higher than for LI_T (Waterson, 1984). See Encaoua and Jacquemin (1980) for a discussion of the properties of proper measures of concentration.

Bibliography

Encaoua, D. and Jacquemin, A. (1980). Degree of monopoly, indices of concentration and threat of entry. *International Economic Review*, **21**, 87 105.

Jacquemin, A. (1987). *The New Industrial Organization: Market Forces and Strategic Behavior*. Cambridge, MA: MIT Press.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Stigler, G. J. (1968). *The Organization of Industry*. Chicago: University of Chicago Press.

Waterson, M. (1984). *Economic Theory of the Industry*. Cambridge: Cambridge University Press.

confidence intervals

Eduardo Ley

In an estimation context, we are often interested in determining a plausible set of values for the unknown parameters rather than to simply rely on single point estimates. Suppose that we want to compute a 95 percent confidence interval, or "an interval with a 95 percent confidence level," for an unknown parameter, β. Then, given our data generation model, we use a rule for constructing an interval for β based on the data such that if the experiment were repeated many times, 95 percent of the time we would expect the interval to indeed include the unknown β. Note that the above is different from the incorrect statement "there is a 95 percent chance that the confidence interval includes the *true* parameter's value." Once the interval is constructed, it either includes the unknown parameter or it does not. All that we know is that if we were to construct such intervals many times with different

samples, we would expect 95 percent of them to include the unknown parameter β.

For instance, suppose that β is a regression coefficient. Then, under fairly general assumptions and when the number of observations is large, the ordinary least squares estimator, $\hat{\beta}$, plus or minus two times its standard error, s, is an approximate 95 percent confidence interval for the unknown coefficient: $\hat{\beta} \pm 2s$. For example, suppose that $\hat{\beta} = 8$ and $s = 2.5$, then the 95 percent confidence interval for $\hat{\beta}$ would be (3,13). Often one divides the OLS estimate over its standard error and compares the result with 2, saying that the estimate is significant (or, significantly different from zero) when it is greater than 2 in absolute value (see T STATIS TIC). What is implicitly being done then is checking whether the 95 percent confidence interval contains zero or not. In the previous example, the interval (3,13) does not include 0, or we could calculate the value of the t statistic for $\hat{\beta} = 8/2.5 = 3.2 > 2$ so the estimate is significantly different from zero (at the 5 percent significance level).

Bibliography

Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.

Silvey, S. D. (1975). *Statistical Inference*. London: Chapman and Hall.

constant cost industry

Robert E. McAuliffe

A constant cost industry is one whose long run supply curve is horizontal at the constant (average) cost of production (see LONG RUN COST CURVES). As the industry expands or contracts, the long run average cost of producing output remains the same. It should be noted that this does not necessarily mean that the firm's production technology exhibits CONSTANT RETURNS TO SCALE; the industry could expand through the ENTRY of additional identical firms with U shaped average cost curves where each firm operates at the minimum point on its AVERAGE TOTAL COST curve.

When demand for a product increases, existing firms produce more output along the short run supply curve (see SHORT RUN COST CURVES) since it is profitable to do so at a higher price. If this increase in demand occurred in a competitive market when the industry was in long run EQUILIBRIUM, the higher price will cause existing firms to earn economic profits since price will equal marginal cost above average total cost (see ECONOMIC PROFIT; MAR GINAL COST). Managers should note that these economic profits will attract entry into the industry in the long run as long as there are no BARRIERS TO ENTRY, and this entry will cause the short run supply curve to eventually shift out and lower the price of the product. Entry will continue until firms in the industry can only earn the normal, risk adjusted rate of return. Therefore in the long run there will be no economic profits earned. If the costs of inputs in production have not been changed by this expansion of the industry, then the new equilibrium price will be the same as before the expansion and the long run supply curve will be flat, connecting these two long run equilibrium prices.

Constant cost industries are more likely to be observed in cases where the inputs in production are not highly specialized or when the industry is small relative to the size of its input markets. These conditions make it more likely that the expansion of the industry (and the consequent increase in the demand for inputs in production) will not cause FACTOR PRICES to rise, and so the minimum average cost remains the same as entry or EXIT occurs in the industry. Managers in constant cost industries can use this knowledge to anticipate price changes. When the industry expands, prices and costs may initially rise, but they will decrease in the long run as entry occurs and prices and costs return to their original values.

See also *economies of scale; perfect competition*

Bibliography

Jehle, G. A. and Reny, P. J. (2000). *Advanced Microeco nomic Theory*, 2nd edn. Boston: Pearson Addison-Wesley.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

constant returns to scale

Robert E. McAuliffe

This describes a specific technology of produc tion for the firm. When constant returns in pro duction exist, a firm that doubled all inputs in production would see a doubling of its output. In the case of the COBB–DOUGLAS PRODUCTION FUNCTION, constant returns to scale occur when the exponents for the inputs in production sum to one. Since with constant returns to scale there are no penalties to the firm whether it expands or contracts, there is no theoretically unique optimal size plant that minimizes the costs of production. With no theoretical limit on the size of firms in the industry, the theoret ical number of firms in the industry is also indeterminate.

See also *economies of scale*

Bibliography

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.
Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.
Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.
Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Publishing.

consumer price index

Robert E. McAuliffe

The consumer price index (CPI) measures the changes in the price of a market basket of goods relative to a base year. The basic question to be answered is: what does it cost in today's prices to buy the same goods purchased in the base year? In the US the Labor Department periodically surveys consumers to determine what they buy and how much. These surveys determine the market basket of goods and the weights given to each price in the index. The last US survey was 1982–4. According to economic theory, the demand for a product depends on the relative (or real) price of the product compared with other goods. Therefore to evaluate the demand curve for their product (*see* DEMAND CURVES), firms should use the CPI to find the relative price of their product, which would be the price the firm charges for its product divided by the CPI. This will give managers a sense of how their product's price compares with the rate of INFLATION and whether their product is becoming relatively more or less expensive versus other goods.

See also *real prices*

Bibliography

Carlson, K. M. (1989). Do price indices tell us about inflation? *Federal Reserve Bank of St. Louis Review*, **71**, 12 30.
Jehle, G. A. and Reny, P. J. (2000). *Advanced Microeco nomic Theory*, 2nd edn. Boston: Pearson Addison-Wesley.
Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

consumer surplus

Robert E. McAuliffe

The difference between the maximum amount a consumer is willing to pay to buy a product (also called the consumer's RESERVATION PRICE) and the price actually paid is called the con sumer's surplus. This measures how much con sumers value a product beyond what they paid for it. In a market, consumer surplus is the area under the demand curve and above the price charged. This measures the benefit to consumers of having this product available and hence the benefits to society from the production of this good if there are no other EXTERNALITIES in production or consumption.

  When a firm raises its price above the MAR GINAL COST, those consumers who continue to buy the product pay a higher price and this represents a reduction in their consumer sur plus. Since these consumers still purchase the

product, it represents a transfer from the con
sumers to the firm. However, some consumers
will not continue to buy the good at the higher
price because the new price exceeds their reser
vation price. Since these consumers were willing
to pay the cost to society of producing the good
(the marginal cost) and do not receive it, they
have lost consumer surplus. Moreover, the firm
does not capture any of this lost surplus as profit
because these consumers have left the market.
Conceptually, the firm should be willing to pro
duce this good and sell it for the marginal cost of
production once the firm has earned its profits in
the market. Since this makes the consumers who
have left the market better off and the firm no
worse off, it is a PARETO OPTIMAL ALLOCA
TION and so allocative EFFICIENCY would be
improved. Because this exchange does not occur,
the consumer surplus of those consumers who
have left the market is lost and this is called
DEADWEIGHT LOSS since no one in society
recovers those lost benefits.

In a similar vein, when a firm considers intro
ducing a new product, it is concerned about
whether it can cover the costs of production,
including the OPPORTUNITY COSTS. If the
firm is unable to cover these costs, the product
will not be produced. Yet it is possible that the
benefits to society from producing the good (as
measured by consumers' surplus) exceed the
costs. The problem is that the firm cannot cap
ture all the benefits to consumers when it cannot
practice PRICE DISCRIMINATION and this can
lead to less PRODUCT VARIETY than would be
socially optimal.

See also *optimal variety*

Bibliography

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic com
petition and optimum product diversity. *American Eco
nomic Review*, **67**, 240 59.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial
Economics: Economic Tools for Today's Decision Makers*,
4th edn. Upper Saddle River, NJ: Prentice-Hall.

Waterson, M. (1984). *Economic Theory of the Industry*.
Cambridge: Cambridge University Press.

Willig, R. D. (1976). Consumers' surplus without apol
ogy. *American Economic Review*, **66**, 589 97.

contestable markets

Robert E. McAuliffe

As originally presented by Baumol, Panzar, and
Willig (1982), perfectly contestable markets rep
resented an alternative and more general bench
mark for evaluating MARKET STRUCTURE than
PERFECT COMPETITION. In perfectly contest
able markets new firms can enter and exit an
industry costlessly and at no disadvantage rela
tive to existing firms, and this potential compe
tition severely limits the pricing options
of existing firms. Costless ENTRY and EXIT at
no disadvantage imply that factors such as
ASYMMETRIC INFORMATION, SUNK COSTS,
FIRST MOVER ADVANTAGES, and STRA
TEGIC BEHAVIOR by existing firms have no
effect on the entrant's decision to enter the in
dustry. Since entrants are at no disadvantage
relative to incumbent firms, whenever existing
firms charge a price sufficient to generate eco
nomic profits, entry will occur driving economic
profits to zero (*see* ECONOMIC PROFIT). Cost
less exit and entry also mean that the entrant can
enter the market, earn economic profits, and exit
before the incumbent firm can retaliate, some
times referred to as "hit and run" entry. For hit
and run entry to succeed, entrants must be able
to enter the market and earn profits before the
existing firm retaliates. This means that prices
must adjust more slowly than output in the
market.

According to Baumol et al., perfectly contest
able markets result in efficient production even
in those cases where perfect competition might
fail, as in NATURAL MONOPOLY. A natural
monopoly faces ECONOMIES OF SCALE where
the average cost of production is falling
in the range of output demanded by the market.
Since average cost is declining, the marginal cost
curve must lie below it, and this means that the
marginal cost pricing required by perfect com
petition would not allow the firm to cover its
total costs (*see* MARGINAL COST). Perfect con
testability requires that the existing firm(s) be
able to cover total costs without attracting entry,
and so a natural monopoly would be forced by
potential COMPETITION to set its price at its
lowest average cost of production.

Perfect competition requires large numbers of small firms producing homogeneous prod ucts, among other conditions, to achieve EFFICIENCY. Perfectly contestable markets emphasize the importance of *potential* competi tors rather than existing competitors and do not require DIMINISHING RETURNS in produc tion. Another important point Baumol et al. raise is that existing firms must produce effi ciently or they will be vulnerable to hit and run entry by potential entrants. This means that certain inefficiencies attributable to monopoly or OLIGOPOLY such as X EFFICIENCY or inefficient organization cannot occur in per fectly contestable markets since entry will force these firms to become efficient or drive them out of business.

Some researchers have argued that few markets are perfectly contestable and that this analytical approach is less general than might at first appear. For example, Shepherd (1984) ob served that perfect contestability requires *ultra* free entry and no exit barriers (no sunk costs). He asserted this was more restrictive than the requirements of perfect competition because in a sufficiently small interval of time, firms always have sunk costs (see Weitzman, 1983). Further more, for potential entry to discipline existing firms to the extent Baumol et al. claim, entry must not only be ultra free but *total* in that entrants could replace the existing firms. Yet if entrants can replace existing firms, the assump tion employed by Baumol et al. that existing firms do not react to entry is very restrictive and if entrants are small, they cannot discipline existing firms to the extent the theory suggests. For Shepherd, these and other concerns lead him to believe that the traditional industry an alysis of industrial organization with its focus on the internal market rivalry and interactions be tween firms is more appropriate.

Bibliography

Baumol, W. J. (1982). Contestable markets: An uprising in the theory of industry structure. *American Economic Review*, **72**, 1 15.

Baumol, W. J., Panzar, J. C., and Willig, R. D. (1982). *Contestable Markets and the Theory of Industry Struc ture*. San Diego: Harcourt Brace Jovanovich.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Schwartz, M. and Reynolds, R. J. (1983). Contestable markets: An uprising in the theory of industry struc ture: Comment. *American Economic Review*, **73**, 488 90.

Shepherd, W. G. (1984). "Contestability" vs. competi tion. *American Economic Review*, **74**, 572 87.

Weitzman, M. L. (1983). Contestable markets: An upris ing in the theory of industry structure: Comment. *American Economic Review*, **73**, 486 7.

contracts

S. Alan Schlacht

Going back to ancient times, contract law has always been a major part of business. Roscoe Pound, a respected legal writer, noted the im portance of contracts when he wrote, "[t]he social order rests upon the stability and pre dictability of conduct, of which keeping prom ises is a large item." Contract law deals with the formation and keeping of lawful promises. The American legal system is based on the right to enter freely a contract or to not do so. In fact, the US Constitution, Article I, Section 10, pro vides that "no state shall pass any law impairing the obligations of contracts." As long as the parties to a contract have a lawful purpose and the intent to enter the agreement, a court will enforce its terms. Contract law is primarily de termined by state law and varies from state to state. However, certain general principles exist.

A contract is generally defined as an agree ment between two or more persons or legal en tities regarding the performance or promise to act or the refraining from acting. A contract is determined from the outward manifestation of intent of the parties. Words and actions of the parties as seen from an objective view determine if a contract exists. When the parties have ex pressly agreed, either verbally or in writing, then the agreement is considered to be an **express contract**. When the parties have shown their intent to be bound by their actions, then the result is an **implied contract**. Courts have even allowed parties to recover compensation

when, although a contract does not exist, one of the parties would be unjustly enriched if no compensation was allowed. This principle is known as quasi contract or **quantum meruit**.

Contracts can also be classified as bilateral or unilateral. A **bilateral contract** involves the promises of one party (the promisor) in exchange for the promises of another (the promisee). These contracts relate to future performance. A **unilateral contract** involves promises in exchange for performance, or vice versa. Thus, some performance is expected for the promise.

Certain elements are required of all contracts in order for a court to enforce the agreement. The first is the meeting of the minds. This element is the **offer** by one party (the offeror) to enter a contract with a corresponding acceptance by the other party (the offeree). An offer is the manifestation of intent by the offeror to enter a binding contact. Expressions of opinion, price quotations, and advertisements are generally considered to be negotiations leading to an offer. In addition, the offer must be reasonably definite. Courts will not supply essential terms that the parties have not agreed upon. Essential terms include at least a description of the subject matter, identification of the parties, the consideration, and in certain situations quantity, location, and time of performance. The offer must also be communicated from the offeror to the offeree in order for a valid acceptance to exist.

An offer may be terminated before acceptance without liability to either party. The death or incompetence of either the offeror or offeree terminates the offer. If the basis of the offer becomes illegal or impossible to perform, then the offer is at an end. The parties may terminate the offer themselves by either revoking the offer (offeror) or rejecting an offer (offeree). Suggesting new terms to an offer by the offeree is a counter offer and ends the original offer. Lastly, an offer may terminate due to the expiration of time. If no time is stated in the offer, then it expires after a reasonable amount of time.

A contract is formed when the offeree accepts the offer. Acceptance must be a mirror image of the offer and must be by some positive act or words of the offeree showing an intent to enter a contract. The acceptance generally must be communicated to the offeror.

The second element of a contract is **consideration**. Consideration is the promise of performance that flows between the parties to a contract. It is also called **legal detriment**. Courts do not require that the exchange between the parties be of comparable value. They only require that something be given up by all parties. Promises made by a party that are not supported by consideration are not enforceable. Thus, a promise to make a gift is generally not enforceable. Consideration states that each party to a contract promise or do an act not otherwise required or refrain from doing an act that is lawfully allowed. The consideration must be bargained for, i.e., present at the time the meeting of the minds occurs. Having performed an act in the past or already being legally obligated to perform an act will not constitute consideration for a current promise. A preexisting duty by law or contract cannot be consideration for a new contract. The only exception is if a promise is made that causes someone to rely on the promise to his detriment. Courts have allowed recovery in this regard under a theory called **promissory estoppel** or detrimental reliance. Moral obligation, however, is not considered to be valid consideration for a contract.

The third element of a contract is **legal capacity**. Courts will not enforce contracts where one or both of the parties do not have the mental or legal competence necessary to engage in such a transaction. The law deems children incompetent to enter contracts due to their age. Thus, minors (generally considered to be persons under the age of 18) have a right to disaffirm a contract for any reason. Once the minor disaffirms, the other contractual party must return all consideration to the minor as long as the minor did not misrepresent his age. Minors cannot disaffirm certain contracts for necessaries, such as medical care or education.

Legal capacity is also lacking when a party has been adjudged incompetent by a court. No contract made by that person thereafter will be enforced. Persons who are intoxicated through alcohol or drug use or persons who are suffering a mental disability are bound to contracts they enter unless they can prove that at the time of contracting they were unable to comprehend the nature and extent of the dealings at hand. If the evidence shows that they could not understand

the legal consequences of their actions, then they are released from the contract.

The final element of a contract is **legality**. All contracts must have a lawful purpose. If a contract would violate a criminal or civil statute, then it is void and no court will entertain a case involving the contract. Examples include most gambling agreements and contracts to harm a person or property. Courts have also fashioned common law illegality by determining that certain contracts are bad for society, i.e., violate public policy. Each of these types of contracts must be examined individually to determine if public policy would be offended if the contract was upheld. Examples include non compete clauses in employment or sale of business contracts, exculpatory clauses, and adhesion contracts. If reasonable under the circumstances, these will be enforced.

In addition to the above elements, certain contracts have to meet further requirements. Under a law passed in England in 1677 and adopted by every state, certain contracts must be in writing to be enforceable. Originally entitled ''A Statute to Prevent Frauds and Perjuries,'' this law is now known as the Statute of Frauds. It requires the following types of contracts to be in writing:

1   any contract that by its terms takes longer than one year to perform (e.g., a two year lease);
2   any agreement relating to the ownership of or liens in real estate;
3   any promise to pay the debt of another (guarantor), an executor's or administrator's promise to pay the debt of the deceased from personal funds; and
4   any promise upon the condition of marriage (e.g., prenuptial agreement).

The Uniform Commercial Code has added certain investment securities and any contract for the sale of goods for $500 or more. Without a writing signed by the other party and containing all the essential terms mentioned above, an aggrieved party is without a remedy. If a contract is found lacking, courts have been willing to award quantum meruit to avoid unjust enrichment. In cases where one party has substantially performed an oral contract, a court may go ahead and enforce the entire contract based on the substantial performance being a substitute for the writing.

Although few contracts are required to be in writing, many are reduced to a writing. Once a contract is put in writing and the document appears to be the entire intent of the parties, no prior or contemporaneous oral statements may be used to vary or contradict the written agreement. Known as the **parol evidence rule**, this principle preserves the integrity of the writing by not allowing a party to attack it using oral evidence. Exceptions do exist, though, including oral evidence to show fraud, mistake, duress, or other reasons to examine the agreement. Also, this rule does not prohibit a subsequent modification of the contract being enforceable.

As discussed above, a contract must be based upon a meeting of the minds of the parties. However, the meeting of the minds can occur due to a lack of genuine assent. This may happen when one or both of the parties is mistaken as to a material fact in the contract. The mistake must be based on fact, not opinion, prediction, or statements of value. A mutual (or bilateral) mistake of fact occurs when both of the parties are mistaken. In this case, a court will rescind the contract and return the parties to their original position. When only one of the parties is mistaken (unilateral), a court will enforce the contract. An exception to this rule is if the other party knows that a mistake has occurred and takes advantage of the mistaken party. A mistake of law is generally not a basis for rescission since everyone is presumed to know the law.

Duress also removes a party's genuineness of assent. Duress is the use of force or fear of personal harm to induce a party to enter a contract. Threatening to kill or harm a person is duress; threatening to fire an employee if she does not enter a contract is *not* duress. In addition, undue influence cancels a party's assent.

Undue influence requires a party in a fiduciary position who takes advantage of a much weaker party for gain. A classic example is a person who induces an elderly relative in his care to deed property to him. Finally, fraud (also called misrepresentation) will nullify the assent needed to form a contract. Fraud is difficult to prove and requires a party to prove five elements:

1   a material misrepresentation of fact exists;
2   it is known to be false;
3   it is done with the intent to deceive;
4   the innocent party justifiably relies on the misrepresentation; and
5   damage occurs.

Courts are reluctant to allow a party to rescind a contract based on fraud if the victim could have exercised reasonable care to avoid the fraud. This defense is known as **caveat emptor**.

Third persons who are not a party to a con tract may enforce a contract if the contract spe cifically contemplates bestowing a benefit on them. An example of such third party benefi ciary would be the beneficiary of a life insurance policy. Other parties may have rights in a con tract if the contract is assigned to them. Any contract may be assigned unless the original agreement prohibits assignment or personal ser vices are involved that require the skill or experi ence of the original party only.

A party can be discharged from further per formance on a contract in several different ways. First, the party can complete her performance. Certain contracts require total performance, such as payment of money or performance by a specific time when time is of the essence. Other contracts are deemed complete even though some performance is still lacking, such as con struction contracts or performance within a rea sonable time. Second, a material breach of contract by the other party relieves a person of continuing to perform, e.g., nonpayment of rent allows the landlord to cancel a lease and evict the tenant. Third, the occurrence of a condition subsequent in the contract, such as a clause that cancels the contract in the event of war or bankruptcy. Fourth, the failure of a condition precedent, such as the failure of a prospective homebuyer to obtain financing, rendering the contract at an end. Fifth, impossibility of per formance, such as the destruction of the subject matter or the death or incompetence of a party. Sixth, the parties may agree to end the contract through rescission or settlement. Finally, a party may be discharged due to bankruptcy or the expiration of the statute of limitations.

If a breach of contract occurs, courts will allow the aggrieved party to recover compensatory damages, i.e., the money necessary to put the party in the position she would be in if the contract had been performed as agreed. Courts rarely allow the recovery of attorney fees unless the agreement so stipulated. Virtually no court will allow punitive damages in a contract action. Late fees and other reasonable charges for minor breaches are not considered punitive and are enforceable. When unique goods or real estate are involved in a breach of contract such that money will not adequately compensate a party, a court may grant a decree of specific performance and order the breaching party to perform. In cases where a court can reform a contract to meet the true intent of the parties, it will do so.

Bibliography

Chirelstein, M. A. (2001). *Concepts and Case Analysis in the Law of Contracts*, 4th edn. New York: Foundation Press.

Murphy, E. J., Speidel, R. E., and Ayres, I. (2003). *Studies in Contract Law*, 6th edn. New York: Foundation Press.

Pound, R. (1959). *Jurisprudence*, vol. 3. Boston: West Publishing Company.

Weaver, J. H. (1990). *The Compact Guide to Contract Law*. Boston: West Publishing Company.

cost of capital

Vickie L. Bajtelsmit

Firms finance their assets and operations with a variety of types of financial CAPITAL, including debt, preferred stock, and common equity. The providers of these funds would not be willing to invest in the firm unless they expected that their investment would provide them with a return sufficient to offset the risk that they have as sumed. Due to the greater certainty of promised cash flows for debtholders (*see* CASH FLOW), the cost of debt is generally lower than that of equity. Similarly, preferred stockholders require lower rates of return than common stockholders since their dividends have priority over common. Costs of all forms of capital will increase with increasing riskiness of the firm.

In making CAPITAL BUDGETING decisions, firms estimate the NET PRESENT VALUE (NPV) of projects by discounting future expected cash flows by an appropriate cost of capital to account for the required returns of

their investors. Thus, the higher the costs associated with investment capital, the harder it is for firms to find acceptable projects. In general, firms employ several types of capital in their capital structure (*see* FIRM FINANCIAL STRUCTURE). Since these funds are not specifically allocated to particular projects or assets of the firm, the appropriate cost of capital to be used in capital budgeting decisions is the weighted average of the various types of capital used.

The procedure used in developing this weighted average is to:

1 estimate the marginal after tax costs of each type of capital; and
2 calculate the weighted average cost of capital (WACC) by weighting each component cost by the percentage of that type of capital in the firm's target financial structure.

In many cases, the firm is subject to limitations on the amount of capital that it can raise at a particular cost level. For example, retained earnings are a less costly form of equity financing since the firm avoids the TRANSACTIONS COSTS associated with sales of newly issued securities. However, retained earnings are limited to the dollars remaining from net income after dividends are paid. Similarly, since risk increases with additional debt, the marginal cost of debt financing will increase as new debt obligations are taken on by the firm.

As an example, suppose that a firm estimates that its marginal after tax cost of new long term debt is 6 percent, the marginal cost of retained earnings is 10 percent, and the marginal cost of newly issued common stock is 12 percent. If they have no preferred stock in their capital structure and the target capital structure calls for 50 per cent debt and 50 percent equity, then the WACC using retained earnings for the equity component is 8 percent, calculated as $0.5 \times$ (6 percent) $+ 0.5 \times$ (10 percent). However, the firm does not have unlimited retained earnings and may be forced to issue common stock to support a larger capital budget. In that case, the marginal cost of capital, using common stock as the equity component, will be 9 percent. Thus, the marginal cost of capital, defined as the cost of the last dollar of new capital the firm raises in a given period, is an increasing function of new dollars

raised in a given period, and the appropriate discount rate to be used for NPV analysis will be the WACC that is applicable for the size of capital budget under consideration.

Bibliography

Brigham, E. F. and Gapenski, L. C. (1997). *Intermediate Financial Management*, 5th edn. New York: International Thomson Publishing.

Gitman, L. J. (1995). *Foundations of Managerial Finance*, 4th edn. New York: HarperCollins.

Cournot competition

Nikolaos Vettas

The term Cournot competition refers to a model of OLIGOPOLY which was introduced by the French mathematician and economist Antoine Augustin Cournot in 1838 and is widely used in industrial organization economics. The model considers a fixed number of firms that choose their output levels in an attempt to maximize their profits. Given the total output produced, the market demand curve determines the price and the profit for each firm (*see* DEMAND CURVES). In the **Cournot equilibrium** each firm does not want to change its output level given the output of its competitors (*see* EQUILIBRIUM). The firms have **Cournot conjectures**, that is, they take the output level of their competitors as given. It is important to recognize that the Cournot equilibrium is just the **Nash equilibrium** of the corresponding simultaneous move game (*see* GAME THEORY). In a Cournot equilibrium the prices are higher (respectively, lower), the total output is lower (higher), and the per firm profit is higher (lower) than in PERFECT COMPETITION (MONOPOLY). In fact, as the number of firms increases, the Cournot equilibrium approaches the perfectly competitive outcome.

The Cournot model has to be distinguished from other models of oligopoly competition such as Stackelberg competition (with FIRST MOVER ADVANTAGES instead of simultaneous moves), Bertrand competition (where firms compete with prices instead of quantities as the strategic variable), and Edgeworth competition (with capacity constraints).

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Industrial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Cournot, A. A. (1927). *Researches into the Mathematical Principles of the Theory of Wealth*, 2nd edn. New York: Macmillan. (Originally published in French as *Recherches sur les principes mathématiques de la théorie des richesses*, 1838.)

Daughety, A. F. (ed.) (1988). *Cournot Oligopoly*. Cambridge: Cambridge University Press.

Vives, X. (1989). Cournot and the oligopoly problem. *European Economic Review*, **33**, 503 14.

covariance

Robert E. McAuliffe

The covariance is a statistical measure of the linear association between two variables, say X and Y, and is denoted as $\text{cov}(X, Y)$. For T observations on variables X and Y, the sample covariance is calculated as

$$\text{cov}(X, \ Y) = \frac{1}{T-1} \sum_{i \ 1}^{T} (X_i - \bar{X}) \times (Y_i - \bar{Y})$$

where $\bar{X}$ is the sample mean for X and $\bar{Y}$ is the sample mean for Y. If there is a positive linear relation between X and Y, the covariance will also be positive, and if there is a negative relation between X and Y, covariance will be negative. When there is no linear relationship between the two variables, the covariance will be zero.

Since the covariance measures the extent to which two variables are related, managers can use this information to determine, for example, how sales are related to a new ADVERTISING campaign. To determine how much sales will change in response to advertising, a LINEAR REGRESSION should be used. The covariance is a component of the estimated regression coefficient (see Berndt, 1991; Maddala, 2001).

Bibliography

Berndt, E. R. (1991). *The Practice of Econometrics*. Reading, MA: Addison-Wesley.

Greene, W. H. (2002). *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.

Gujarati, D. N. (2002). *Basic Econometrics*, 4th edn. New York: McGraw-Hill Irwin.

Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.

credence goods

Robert E. McAuliffe

Some EXPERIENCE GOODS are difficult to evaluate before and after purchase and these goods are called credence goods. Medical services, automotive repairs, and investment advice are examples of goods where consumers are unlikely to be able to judge product quality. In such cases it is expensive for consumers to acquire enough information to determine product quality. Unfortunately firms have incentives to take advantage of consumers' lack of information and provide a low quality good at a high price. However, established firms have incentives to invest in BRAND NAME capital to help reduce these information costs.

See also *asymmetric information; quality*

Bibliography

Darby, M. R. and Karni, E. (1973). Free competition and the optimal amount of fraud. *Journal of Law and Economics*, **16**, 67 88.

Shughart, W. F., Chappell, W. F., and Cottle, R. L. (1994). *Modern Managerial Economics*. Cincinnati, OH: South-Western Publishing.

credible strategies

Robert E. McAuliffe

Existing firms may try to prevent potential ENTRY by using strategic threats. But for these threats to be believed, they must be the profit maximizing choice the incumbent firm would make should entry occur. For example, an established firm might threaten to lower prices if another firm enters the market. This threat will not be credible, however, if by lowering prices the existing firm were made worse off. Therefore, for a threat or strategy to be credible, it must be in the best interests of the established firm in the post entry period. Usually firms can

make their threats credible by making commit ments in previous periods (*see* COMMITMENT). For example, a firm could invest in additional capacity to make a threat to lower prices in subsequent periods credible.

When entrants do not have perfect informa tion about the established firms, it is possible for established firms to bluff and make threats which would not maximize their profits in the post entry period. As long as potential entrants believe the threat could occur with a sufficiently high probability, entry can be deterred (see Encaoua, Geroski, and Jacquemin, 1986). Cred ible threats represent one element of STRA TEGIC BEHAVIOR that incumbent firms can undertake to affect their MARKET STRUCTURE.

See also *accommodation; first mover advantages*

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Indus trial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Encaoua, D., Geroski, P., and Jacquemin, A. (1986). Strategic competition and the persistence of dominant firms: A survey. In J. E. Stiglitz and G. F. Mathewson (eds.), *New Developments in the Analysis of Market Structure*. Cambridge, MA: MIT Press.

Jacquemin, A. (1987). *The New Industrial Organization: Market Forces and Strategic Behavior*. Cambridge, MA: MIT Press.

Salop, S. C. (1979). Strategic entry deterrence. *American Economic Review*, **69**, 335 8.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

cross-advertising elasticity

Robert E. McAuliffe

The cross advertising elasticity of demand measures the percentage change in the quantity sold of one product given a percentage change in the ADVERTISING for another product. For products which are COMPLEMENTS, an in crease in advertising for, say, automobiles may also increase the quantity of gasoline sold, and so the cross advertising elasticity would be posi tive. Products which are SUBSTITUTES would have a negative cross advertising elasticity since the advertising of one automotive company, Ford, should have a negative effect on the quan tity of automobiles sold by General Motors, all else equal. For a company with a product line, the cross advertising elasticity may also be im portant since increased advertising of one brand in the product line could reduce sales of a closely related brand.

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Indus trial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

cross elasticities

Gilbert Becker

The quantity demanded (or supplied) of a good depends not only on its own price but on the prices of other goods as well (*see* DEMAND FUNCTION). The cross elasticity of demand is the measure of the effect on the quantity demanded of one product (x) resulting from the change in the price of another (y). It is measured as

$$\varepsilon_{xy} = \frac{\Delta Q_x}{Q_x} \Big/ \frac{\Delta P_y}{P_y} = \frac{\Delta Q_x}{\Delta P_y} \times \frac{P_y}{Q_x}$$

where the symbol Δ indicates a change in quan tity (Q) or price (P) and the subscripts (x,y) identify the two different goods. For goods which are SUBSTITUTES, such as imported and domestic beer, the cross elasticity value will be positive since an increase in the price of one good will increase the demand for the other. Goods which are COMPLEMENTS, such as hamburgers and hamburger rolls, will have a negative value for cross elasticity. The measure ment of cross elasticity is useful to a firm in that it indicates the extent to which consumers perceive a rival product as being a close substi tute. A rise in the price of one's own product under circumstances where the cross elasticity with other products is high can lead to signifi cantly diminished unit sales and revenues. The

measure is also useful in developing an economic definition of the boundaries of a market (*see* MARKET DEFINITION) but may be of limited practical value in antitrust analysis.

See also *complements; substitutes*

Bibliography

Browning, E. and Zupan, M. (2004). *Microeconomics: Theory and Applications*, 8th edn. New York: John Wiley.
Maurice, S. and Thomas, C. (2002). *Managerial Economics*, 7th edn. Boston: McGraw-Hill Irwin.
Miller, R. L. (2004). *Economics Today: The Micro View*, 12th edn. Boston: Pearson.

cross-section analysis

Robert E. McAuliffe

Data which is gathered over a number of house holds or firms at the same point in time is called cross section data. Managers can use this source of quantitative information for LINEAR RE GRESSION to determine, for example, whether there are ECONOMIES OF SCALE in produc tion. In this case, managers would collect cost data from firms producing the same product at the same point in time, but with different cap acities. Economies of scale would exist if firms with larger capacities had lower average total costs of production (*see* AVERAGE TOTAL COST). The econometric tools required for cross section analysis are sometimes different than those required for time series analysis (*see* TIME SERIES DATA; TIME SERIES FORE CASTING MODELS). For example, HETERO SKEDASTICITY arises more frequently in cross section data than in time series (see Greene, 2002; Gujarati, 2002; Maddala, 2001).

Bibliography

Greene, W. H. (2002). *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.
Gujarati, D. N. (2002). *Basic Econometrics*, 4th edn. New York: McGraw-Hill Irwin.
Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.

D

deadweight loss

Wei Li

The term "deadweight loss" refers to the welfare loss due to an inefficient allocation of resources in which the market price of a good deviates from the MARGINAL COST of producing that good. Inefficient allocations of resources can occur in a variety of situations including, for example, the situation where a market is not perfectly competitive or the situation where a sales tax is imposed on a perfectly competitive industry (*see* EFFICIENCY; PERFECT COMPETITION).

When resources are allocated inefficiently (e.g., as a result of a sales tax), there are typically two effects:

1 welfare transfers that simply redistribute CONSUMER SURPLUS and PRODUCER SURPLUS among buyers, sellers, and any other party (e.g., the government); and
2 net reductions of consumer surplus and/or producer surplus that represent welfare loss to the economy.

For example, a higher market price due to a tax will cause some consumers and/or producers to leave the market. The government does not receive any tax revenues from those consumers or producers who no longer participate in the market but who are now worse off. This welfare loss is not offset by any gain to either the sellers, the buyers, or any other party. It is hence named deadweight loss.

Bibliography

Bergson, A. (1973). On monopoly welfare losses. *American Economic Review*, **63**, 853 70.
Harberger, A. (1964). Taxation, resource allocation and welfare. In J. Due (ed.), *The Role of Direct and Indirect Taxes in the Federal Revenue System*. Princeton, NJ: Princeton University Press.
Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

declining industry

Robert E. McAuliffe

Typically a declining industry is a mature industry whose sales are decreasing. Such an industry is in the final stage of its life cycle, although industry sales may revive at some point (Jacquemin, 1987; Porter, 1980; *see* INDUSTRY LIFE CYCLE; PRODUCT LIFE CYCLE). When an industry's sales are declining, net rates of EXIT by firms will increase either through failure or through acquisition by other firms. The industry structure may consolidate and industry concentration may rise. In this stage of the industry life cycle, surviving firms often must compete on the basis of costs since many SUBSTITUTES are available to consumers.

If there are exit barriers or if firms are unwilling to leave the industry, competition between the firms that remain in a declining industry may intensify. In the extreme case, firms may engage in a "war of attrition" where no firm wants to be the first to leave the industry since the remaining firm or firms will be able to earn economic profits (*see* ECONOMIC PROFIT). Porter (1980) identifies several sources of exit barriers including specialized assets (*see* ASSET SPECIFICITY), fixed costs of exit (such as settling labor and other CONTRACTS), and strategic barriers where presence in the declining market may

confer advantages for the firm in other markets in which it sells. Whinston (1988) theoretically evaluated the social welfare consequences of exit and observed that firm exits from a declining industry may not be optimal since the ''wrong'' (lower cost) firm may exit before the higher cost firm. Furthermore, multiplant firms may reduce their capacity when the efficient outcome would require that small firms exit the industry. These welfare effects are similar to the biases from free ENTRY but in reverse (see also Tirole, 1988).

Bibliography

Jacquemin, A. (1987). *The New Industrial Organization: Market Forces and Strategic Behavior*. Cambridge, MA: MIT Press.

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

Whinston, M. D. (1988). Exit with multiplant firms. *The Rand Journal of Economics*, **19**, 568 88.

decreasing cost industry

Robert E. McAuliffe

A decreasing cost industry is one where costs decrease as the industry expands. In this case the industry's long run supply curve slopes downward: as the industry produces more output, the minimum average cost of production for each firm decreases with the decrease in costs (*see* LONG RUN COST CURVES). Firms in a decreasing cost industry do not necessarily have ECONOMIES OF SCALE in production; the decrease in costs may reflect lower input costs which reduce the minimum point of the AVERAGE TOTAL COST curve as the industry grows.

Input costs may decline as the industry expands if there are economies of scale in the production of an important input. For example, economies of scale in the production of computer chips allow personal computer manufacturers to produce more computers at lower cost as chip prices fall. An industry may also experience decreasing cost if there are ''economies of agglomeration.'' These economies can occur when a number of firms produce in a specific geographic area and, as their number grows, supporting services such as transportation can be provided to all firms at lower cost. Again, this lowers each firm's costs as the industry grows and so the minimum point on the average cost curve where there are zero economic profits is lower (*see* ECONOMIC PROFIT). Managers in decreasing cost industries can use this knowledge to anticipate price changes. When the industry expands due to an increase in demand, prices and costs may initially rise but they will decrease *below* their original levels in the long run as ENTRY and industry output increase. Decreasing cost industries have long run supply curves which are negatively sloped.

Bibliography

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Martin, S. (2001). *Advanced Industrial Economics*, 2nd edn. Oxford: Blackwell.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

demand curves

Robert E. McAuliffe

For individuals or markets, demand curves show the relationship between the quantity demanded and the price of the good (holding everything else constant). Given the LAW OF DEMAND, the quantity demanded will be inversely related to the relative price of the good. The demand curve will shift whenever any other variables in the DEMAND FUNCTION change (except the price of the product itself). Since TOTAL REVENUE is equal to price, P, times quantity, Q, and the demand curve expresses quantity demanded as a function of price alone, the demand curve is also the average revenue curve. In other words, dividing total revenues (sales) by the number of units sold yields the average revenue (price) per unit sold.

Bibliography

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Martin, S. (2001). *Advanced Industrial Economics*, 2nd edn. Oxford: Blackwell.

Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Publishing.

demand function

Robert E. McAuliffe

The demand function shows the relationship between the quantity demanded and *all* vari ables or factors which affect demand. These variables include the price of the product itself, the prices of related goods (SUBSTITUTES and COMPLEMENTS), consumer income levels, consumer preferences, the information available to consumers, the product's quality, consumer expectations, and the ADVERTISING and pro motional efforts for the product and for com peting products. Additional variables may include the population in the market, the weather, and other factors specific to a prod uct's market.

Of the variables listed above, firm managers can control only a few. A firm can exert some control over the price it charges, its own adver tising outlays, and the quality of its product (including the services offered before and after the sale). Douglas (1992) refers to these variables as strategic variables which the firm can use to enhance its market position.

The specific form the demand function will take depends on the relationship between the quantity demanded and these variables in a given market and time period. For example, the Cobb–Douglas form of the demand function allows the variables to affect demand multiplica tively; in the simplest case where only three variables affect the demand for the product, the Cobb–Douglas demand function is:

$$Q_x^d = P_x^\alpha A_x^\beta I_x^\gamma$$

where P_x is the price of product x, A_x is the level of advertising for the product, and I_x is the real income of consumers in the market. In this example, the coefficients α, β, and γ would measure the ELASTICITY of demand with re spect to its own price, advertising, and income, respectively. Other functional relationships for demand may be specified and estimated (Berndt, 1991; *see* LINEAR REGRESSION).

Bibliography

Berndt, E. R. (1991). *The Practice of Econometrics*. Read ing, MA: Addison-Wesley.

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Martin, S. (2001). *Advanced Industrial Economics*, 2nd edn. Oxford: Blackwell.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Publishing.

depreciation

Laura Power

Depreciation is the decrease in worth of a durable good (often a capital asset used in pro duction) which is associated with its use or the passage of time. It is a component of the cost of production for businesses and therefore must be included in the calculation of ECONOMIC PROFIT.

Although a capital good is purchased at a single point in time, its use occurs slowly over time. Therefore, some structure must be im posed to determine the decline in the asset's value over time; that is, the rate at which the asset is depreciated. The two most common measurement methods are **straight line** and **declining balance**. Both distribute the depre ciation incrementally over the useful life of the asset. However, the straight line method depre ciates it in equal increments over the useful life, whereas the declining balance method depreci ates it more rapidly in the initial years after purchase.

Specifically, for an asset which is purchased in period t and has a useful life of T years, $t = 1$ to T, the straight line method determines depreci ation as:

$$D(t) = B/T$$

where $D(t)$ is the depreciation for period t, B is the initial value of the asset minus the scrap value of the asset. For the declining balance method of depreciation, the relationship is:

$$D(t) = B(t-1) \times R$$
$$B(t) = (B(t-1) - D(t))$$
$$R = DBR/T$$

where $D(t)$ is the depreciation for year t, $B(0)$ is the initial value of the asset minus its scrap value, $B(t)$ is the basis value for t, and DBR is the chosen rate of depreciation for the declining balance.

Because depreciation assigns a stream of expenditures to the purchase of a capital asset, it has been used as the value of the tax deductions which have historically been offered for investment. The method applicable for tax purposes is assigned by law; therefore, it can change over time, and is not necessarily identical to the accounting method chosen by a given business. The *1993 Depreciation Guide* provides a summary of the measurement methods, assignment of class lives, and history of the depreciation methods utilized for tax purposes. In economic literature, depreciation studies have ranged from vintage capital models (see, e.g., Cooper and Haltiwanger, 1993) to estimates of economic depreciation rates.

Bibliography

CCH Tax Law Editors (1993). *1993 Depreciation Guide featuring MACRS*. Chicago: Commerce Clearing House.

Cooper, R. and Haltiwanger, J. (1993). The macroeconomic implications of machine replacement: Theory and evidence. *American Economic Review*, 83, 360 82.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Publishing.

diminishing marginal utility

Robert E. McAuliffe

As consumption of any good increases within a specific period of time, additional units of the good will yield less satisfaction (or TOTAL UTILITY). That is, as a person consumes additional glasses of, say, soda, the additional satisfaction from consuming more soda (the MARGINAL UTILITY) will eventually decline. Diminishing marginal utility had been assumed to obtain INDIFFERENCE CURVES that were convex to the origin and to obtain downward sloping demand curves from consumer choice. However, it is not clear that there is diminishing marginal utility for income (the consumption of all goods), so the less restrictive assumption of a diminishing MARGINAL RATE OF SUBSTITUTION between any two goods is now employed (see Douglas, 1992; Keat and Young, 2002; Stigler, 1966).

Bibliography

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Stigler, G. J. (1966). *The Theory of Price*, 3rd edn. New York: Macmillan.

diminishing returns

Robert E. McAuliffe

When additional units of a factor of production are added to fixed amounts of other inputs in production, at some point the increase in output which results will decrease. Holding other inputs constant means that the level of technology used to combine inputs in production is also held constant. There may be diminishing total returns where total output actually falls from additional units of an input – as when too many workers in a plant reduce the total output produced by interfering with each other, in which case the TOTAL PRODUCT of labor would be declining. Or there may be diminishing average returns where the average output increases by less from additional units of an input. If the input which is increased is labor, the average product of labor would be decreasing in this case.

Diminishing *marginal* returns occur when additional units of an input result in a smaller

increase in output (or, equivalently, the MAR GINAL PRODUCT of that input will decline). Since diminishing returns applies when at least one input in production is held constant, it rep resents a SHORT RUN phenomenon. When all inputs in production are varied, the question is one of returns to scale.

See also *economies of scale; increasing returns; law of variable proportions*

Bibliography

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.
Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.
Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Publishing.
Stigler, G. J. (1966). *The Theory of Price*, 3rd edn. New York: Macmillan.

components very efficiently. As this occurred, changes in technology and factor costs decreased the MINIMUM EFFICIENT SCALE for computer manufacturing. Even if IBM's separate divisions were as efficient as their competitors, the in ternal pricing and cost information did not re flect this efficiency. The organization has been restructured so that the divisions operate as sep arate companies to be closer to the market (*see* VERTICAL INTEGRATION).

Bibliography

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.
Martin, S. (2001). *Advanced Industrial Economics*, 2nd edn. Oxford: Blackwell.
Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.
Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Publishing.

diseconomies of scale

Robert E. McAuliffe

Diseconomies of scale are said to exist when long run average total costs increase because a firm is too large (*see* AVERAGE TOTAL COST). These diseconomies may rise because of technological factors but are usually attributed to decreasing returns to management. As the size of the firm increases, the firm's planning and coordinating activities become more unwieldy and as more bureaucratic layers are added to the organiza tion, managers are separated from the market and their customers. If consumer preferences change rapidly or if new products are frequently introduced by rivals, larger firms may be disad vantaged by their lack of flexibility and slow adjustment to the market.

An interesting example has been the reorgan ization of IBM. In the early years of the com puter industry, there were no independent suppliers of computer chips and peripheral devices and computer manufacturers had to be vertically integrated to produce those inputs. But as the industry grew, so did independent suppliers who could specialize and produce

dominant firm

Robert E. McAuliffe

A dominant firm has significant MARKET POWER which it can exercise over other firms in the industry (often called fringe firms or the competitive fringe). The dominant firm is so much larger than the fringe firms that it does not need to consider their reactions to its deci sions (unlike an OLIGOPOLY market structure). It maximizes its profits according to its RE SIDUAL DEMAND and the price taking fringe firms produce along their supply curve at that price. Stigler (1968) suggested a market share of 40 percent or more was required for a firm to be considered dominant, and if the second firm in the industry was large, the required market share would be even higher. Other economists have suggested market share thresholds between 30 and 60 percent (see Greer, 1992). The dominant firm model may also apply when a group of firms form a cartel in an industry and act as a dominant firm toward non member firms (*see* CARTELS).

In a single period setting where all firms sell identical products, the dominant firm must con sider the output which will be supplied by the

fringe firms in order to maximize its profits. Attempting to raise price by restricting output will cause fringe firms to produce more output, and this will reduce the industry price. There fore the dominant firm will choose its output level so that, given the output of the fringe firms at each price, profits are maximized. That is, the dominant firm should subtract the supply of the fringe firms from market demand and maximize profits by producing until MAR GINAL REVENUE equals MARGINAL COST along this new (residual) demand curve (*see* DEMAND CURVES). This will determine the profit maximizing output and the industry price, and will also determine the fringe firms' output. As long as the smaller firms charge a price equal to or below that of the dominant firm, they will sell their output. The dominant firm is sometimes said to provide a *price umbrella* under which less efficient firms can survive, but this will only be true if there are entry barriers which prevent efficient firms from entering the competitive fringe (Stigler, 1968; *see* BARRIERS TO ENTRY). Dominant firm pricing is also referred to as a PRICE LEADER SHIP model, although there are other forms of price leadership.

The pricing decisions of a dominant firm are constrained by the presence of the fringe firms, and this prevents the firm from charging a price as high as the MONOPOLY price. A monopolist has greater market power because it is protected by significant barriers to entry. When there are no barriers to protect the dominant firm, if it sets its price too high, then it will encourage expan sion by the fringe firms through new ENTRY, which will reduce prices and will eventually erode the market position of the dominant firm. Gaskins (1971) showed that if the domin ant firm had no cost advantages and set its prices optimally over time, it could slow the rate of market penetration by new entrants but would ultimately lose its dominant position.

Dominant firms may increase their market power through STRATEGIC BEHAVIOR aimed at deterring entry. Instead of relying on price adjustments alone to deter entry (as Gaskins assumed), dominant firms can also increase their ADVERTISING, the number of products they produce, they may develop aggressive repu tations, or maintain EXCESS CAPACITY to threaten potential entrants (see Encaoua, Ger oski, and Jacquemin, 1986). These additional competitive weapons, if successful, would enable dominant firms to maintain their market pos itions for longer periods of time. In his survey of executives at major American corporations, Smiley (1988) found that managers preferred to use PRODUCT PROLIFERATION and high ad vertising to discourage potential entry.

See also *limit pricing; price leadership*

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Indus trial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Encaoua, D., Geroski, P., and Jacquemin, A. (1986). Strategic competition and the persistence of dominant firms: A survey. In J. E. Stiglitz and G. F. Mathewson (eds.), *New Developments in the Analysis of Market Structure*. Cambridge, MA: MIT Press.

Gaskins, D. (1971). Dynamic limit pricing: Optimal pricing under threat of entry. *Journal of Economic Theory*, **2**, 306 22.

Greer, D. F. (1992). *Industrial Organization and Public Policy*, 3rd edn. New York: Macmillan.

Smiley, R. (1988). Empirical evidence on strategic entry deterrence. *International Journal of Industrial Organ ization*, **6**, 167 80.

Stigler, G. J. (1968). *The Organization of Industry*. Chi cago: University of Chicago Press.

Dorfman Steiner condition

Robert E. McAuliffe

A firm can increase its unit sales by decreasing price or by increasing ADVERTISING. For a given price, if advertising increases sales by one unit, the revenue from the increased advertising is the profit on that unit sold, or $(P - MC)$ where P is the price of the product and MC is the MARGINAL COST of producing that unit. All else equal, firms should advertise more the higher the profit margin per unit sold.

Dorfman and Steiner (1954) considered the case where the demand for a monopolist's prod uct is affected by both price and advertising and asked what the optimal (profit maximizing) combination of price adjustments and advertis ing might be. Demand for the product depends

upon (is a function of) price, P, and the number of advertising messages, A:

$$Q^d = f(P, A)$$

Profits for the monopolist will be maximized at the output level where the price–cost margin is equal to the inverse of the absolute value of the price ELASTICITY of demand (*see* MARKUP PRICING). Since advertising can change the price elasticity of demand, the firm will want to advertise until the MARGINAL REVENUE from advertising equals the marginal cost of advertising. The marginal revenue from advertising is the profit per unit (the price–cost margin) while the marginal cost of advertising is the cost of an advertising message, c_a, times the number of advertising messages purchased, A. There fore, if advertising increases the firm's sales by ΔQ, the condition where marginal revenue is equal to marginal cost is:

$$c_a \times \Delta A = \Delta Q \times (P - \text{MC})$$

and this can be rearranged to obtain the Dorf man–Steiner condition:

$$\frac{c_a A}{PQ} = \frac{\varepsilon_A}{-\varepsilon_P}$$

Since total advertising expenditures are in the numerator of the left hand side of the equation above and total sales are in the denominator, the left hand side is the advertising–sales ratio. On the right hand side, ε_A is the elasticity of demand with respect to advertising and it shows the percentage change in unit sales for a given percentage change in advertising mes sages, while the denominator is the negative of the price elasticity of demand.

The Dorfman–Steiner condition represents the profit maximizing level of advertising for a monopolist. When a firm introduces a new prod uct, we would expect demand to be very respon sive to advertising, so ε_A should be high. This means that the advertising–sales ratio will be higher for new products in the early stages of their PRODUCT LIFE CYCLE than for mature products. In a growing market the elasticity of demand with respect to advertising will be higher and so will the advertising–sales ratio, while the advertising elasticity of demand will be lower in declining markets where each firm's

sales will come at the expense of rival firms. The optimal advertising–sales ratio will increase as the (absolute value of the) price elasticity of demand decreases. Firms would rather advertise more rather than lower prices as the price elasti city of demand decreases. Advertising will be higher the higher the profit per unit sold, so an empirical relationship between advertising and profits would be expected given the Dorfman–Steiner result (see Berndt, 1991; McAuliffe, 1987; Schmalensee, 1972). Schmalensee (1972) extends the model to an OLIGOPOLY where firms have Cournot expect ations about how rival firms will react (see COURNOT COMPETITION).

Bibliography

Berndt, E. R. (1991). *The Practice of Econometrics*. Read ing, MA: Addison-Wesley.

Dorfman, R. and Steiner, P. O. (1954). Optimal advertis ing and optimal quality. *American Economic Review*, **44**, 826 36.

McAuliffe, R. E. (1987). *Advertising, Competition, and Public Policy: Theories and New Evidence*. Lexington, MA: D. C. Heath.

Martin, S. (2001). *Advanced Industrial Economics*, 2nd edn. Oxford: Blackwell.

Schmalensee, R. (1972). *The Economics of Advertising*. Amsterdam: North-Holland.

dumping

Kent A. Jones

According to the traditional definition, dumping is the practice of PRICE DISCRIMINATION in international trade, in which the exporter charges a lower price for a specific product in the export market than in his home market. International trade law, as embodied in the Gen eral Agreement on Tariffs and Trade (GATT), Article VI, recognizes two additional definitions of dumping, which can be applied if the export er's home price is deemed inappropriate as a basis for comparison:

1   charging a lower price for a product in one export market than in another export market; and

2   charging a price that does not cover the cost of production, including a "reasonable" add ition for selling cost and profit.

US international trade law generalizes the defin ition of dumping as the sale of an imported product at "less than fair value," according to the applicable basis of price comparison. According to GATT rules, if an investigation finds that dumping has taken place and "in jures" a domestic industry (see below), the importing country can impose an antidumping duty in the amount of the difference between the export price and the "fair value" price.

Viner (1923) was the first to offer a systematic investigation of dumping. For the purposes of economic analysis, the central questions focus on the motivation for and welfare effects of dumping. If an exporting firm with price making power has the ability to isolate markets with differing price elasticities of demand, for example, simple profit maximizing behavior motivates a systematic pricing policy of dumping as an international form of third degree price discrimination ("persistent" dumping, in Viner's terms). Typically, factors such as trans portation cost or import restrictions in the exporter's home country, as well as an inter national MARKET STRUCTURE restricting com petition, contribute to the exporter's ability to price discriminate. In addition, temporary sur pluses may lead to "sporadic" dumping and third party consignment sales may lead to pricing differentials that can be characterized as "inadvertent" dumping. In these scenarios, dumping generally improves consumer welfare in the importing country while decreasing the welfare of import competing producers, with a net gain to the importing country as long as competition itself is not significantly reduced.

The main focus of antidumping laws, how ever, is the fear of PREDATORY DUMPING (cf. PREDATORY PRICING), which is presumably motivated by a strategy by the exporter of under cutting prices of domestic producers in the targeted export market in order to drive them out of business and monopolize the market, thus decreasing total welfare in the importing coun try. Typically, such a strategy would require pricing below the MARGINAL COST, which differs significantly from that of simple price

discrimination. In addition, the "cost of produc tion" definition of dumping, described above, may merely reflect the loss minimizing practice of equating MARGINAL REVENUE and marginal cost and then setting price below AVERAGE TOTAL COST (but above the shutdown point of the firm) when the firm's demand curve lies below its average total cost curve (*see* AVERAGE VARIABLE COSTS; DEMAND CURVES). In short, dumping may merely reflect traditional profit maximizing/loss minimizing behavior by firms in international markets that does not in volve predatory motives.

Although the conditions for a successful preda tory strategy are difficult to fulfill (see Boltuck and Litan, 1991: chap. 1), antidumping laws are driven principally by the fear of predatory dumping, whether or not there is evidence that the exporter is capable of pursuing such a strat egy. According to GATT rules, in order to impose antidumping duties an antidumping in vestigation must establish (1) that dumping has taken place and (2) that the dumping causes or threatens "material" injury to a domestic indus try.

See also *profit maximization*

Bibliography

Boltuck, R. and Litan, R. E. (1991). *Down in the Dumps: Administration of the Unfair Trade Laws*. Washington, DC: Brookings Institution.

Viner, J. (1923). *Dumping: A Problem in International Trade*. New York: Augustus M. Kelley.

Durbin Watson statistic

Robert E. McAuliffe

One of the most common tests for autocorrela tion which is frequently reported in statistical software is the Durbin–Watson statistic. This statistic tests for first order autocorrelation using the estimated residuals from a LINEAR REGRESSION. The statistic is calculated as:

$$DW = \frac{\sum\limits_{t\ 2}^{T}(e_t - e_{t\ 1})^2}{\sum\limits_{t\ 1}^{T}(e_t)^2}$$

where e_t is the estimated residual for observation t and e_{t-1} is the estimated residual for observation $t - 1$. The numerator measures (approximately) the COVARIANCE between successive observations of the estimated residuals and the denominator measures the VARIANCE. Note that in order to calculate the covariance in the numerator, the first observation is lost. It can be shown that the Durbin–Watson statistic will lie between 0 and 4 (see Durbin and Watson, 1951; Gujarati, 2002; Maddala, 2001).

When a positive error tends to be followed by another positive error, the covariance in the numerator will be positive and the Durbin–Watson statistic will have a value less than two. If there is no relationship between successive values of the error terms, then the statistic will have a value equal to two, while a negative relation between successive values of the estimated error terms causes the statistic to have a value above two. Since the sampling distribution of the Durbin–Watson statistic is affected by the explanatory variables included in the regression, special tables are used to test the hypothesis of zero first order autocorrelation. The Durbin–Watson test does have several weaknesses: it only tests for first order autocorrelation, the test can be inconclusive, and the test is biased when there are lagged dependent variables in the regression (see Berndt, 1991; Greene, 2002; Maddala, 2001).

Bibliography

Berndt, E. R. (1991). *The Practice of Econometrics*. Reading, MA: Addison-Wesley.

Durbin, J. and Watson, G. S. (1951). Testing for serial correlation in least squares regression. *Biometrika*, **38**, 159 77.

Greene, W. H. (2002). *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.

Gujarati, D. N. (2002). *Basic Econometrics*, 4th edn. New York: McGraw-Hill Irwin.

Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.

E

economic depreciation

Robert E. McAuliffe

Sometimes referred to as "Hotelling" depreci ation (see Fisher and McGowan, 1983; Edwards, Kay, and Mayer, 1987), economic depreciation reflects the change in the value of an asset during a period. This change in the value of the asset includes physical depreciation and any changes in the market value of the asset. If the PRESENT VALUE of an asset's remaining cash flows (dis counted at the asset's INTERNAL RATE OF RETURN) has changed from the beginning of the period to the end (*see* CASH FLOW), that change represents the economic depreciation. Since accounting standards apply somewhat ar bitrary depreciation schedules to physical assets (such as straight line depreciation, sum of year's digits, and double declining balance), it is unli kely that reported depreciation will be equal to economic depreciation. As a result of this problem and others, ACCOUNTING PROFIT will not equal ECONOMIC PROFIT and the acco unting rate of return will not equal the economic rate of return. This feature is important in economic analysis because it implies that studies which use accounting profit calculations may be biased in their results (see Fisher and McGowan, 1983).

See also *accounting profit; depreciation; economic profit*

Bibliography

Edwards, J., Kay, J., and Mayer, C. (1987). *The Economic Analysis of Accounting Profitability*. Oxford: Oxford University Press.
Fisher, F. M. and McGowan, J. J. (1983). On the misuse of accounting rates of return to infer monopoly profits. *American Economic Review*, **73**, 82 97.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.
Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Publishing.

economic profit

Robert E. McAuliffe

Economic profits are defined as total revenues in a given period minus *all* costs of production in that period, including the economic costs of tangible and intangible assets and the opportun ity cost of those assets employed (*see* OPPOR TUNITY COSTS). For example, any costs incurred for training, research and development, and other intangible assets would be depreciated over their expected economic lifetime rather than expensed in the current period. The normal, risk adjusted rate of return to capital would be included in the calculation of economic costs, so that firms which earned the normal rate of return would have zero economic profits, although they would show positive accounting profits (*see* ACCOUNTING PROFIT). When eco nomic profits are positive, they serve as a signal for ENTRY into the industry, while negative economic profits indicate that firms should EXIT the industry. In the long run, economic profits in competitive markets should be driven to zero by the forces of entry and exit unless there are BARRIERS TO ENTRY or bar riers to exit. In fact, Bain (1941) suggested using the accounting profit rate above the competitive return as a measure of the height of entry barriers in an industry. However, this is an appropriate measure only when the

accounting rate of return equals the economic rate of return.

Fisher and McGowan (1983) argued that the appropriate theoretical measure of the economic rate of return is the INTERNAL RATE OF RETURN (IRR), and they asserted there was little relation between the accounting rate of return and the economic rate of return. The accounting rate of return is calculated as acco unting profits divided by either total assets or owners' equity. But accounting statements report assets at their historic costs and not at their current (or replacement) cost, so inflation and the failure to capitalize intangible asset values will cause accounting rates of return to deviate from the economic rate of return. Fisher and McGowan showed that the accounting rate of return will not equal the economic rate of return because:

1   the DEPRECIATION methods used by accountants do not reflect ECONOMIC DEPRECIATION;
2   the investments made by a firm in any given period will not generally have the same "time shape" (stream of future cash flows; *see* CASH FLOW); and
3   the accounting rate of return will vary with the rate of growth of the firm.

From their simulations using the same stream of cash flows from a hypothetical investment but different methods of depreciation, Fisher and McGowan found that the accounting rate of return was a misleading measure and concluded it may provide no information about the eco nomic rate of return.

Not surprisingly, these results generated con siderable controversy in the literature. Long and Ravenscraft (1984) and Martin (1984) suggested that other measures of profitability, such as the LERNER INDEX, may serve as better measures of MARKET POWER for economic analysis. Long and Ravenscraft also argued that in prac tice, the correlation between accounting profits and economic profits was very high and the examples used by Fisher and McGowan were not representative of industry. Based on these arguments, Long and Ravenscraft concluded ac counting data provide useful information about the economic rate of return. Salamon (1985) also found that there was a strong correlation be tween the accounting rate of return and the estimated *conditional* internal rate of return (IRR) which can be obtained from firms' finan cial statements. However, he also found system atic errors occurred when the accounting rate of return was employed, so the accounting measure could not simply be treated as a randomly "noisy" proxy for the economic rate of return. In particular, he found that the errors from using the accounting rate of return varied systematic ally with firm size and therefore jeopardized conclusions from studies relating accounting profits to industry concentration. The problem arose, in part, because the method of depreci ation firms choose varies systematically with firm size.

Fisher (1984) noted in his reply that the fun damental problem of accounting rate of return data is that profits in the numerator are measured as an average of *past* investments, while total assets in the denominator are based on historic costs and may include recent capital that has been added in the expectation of future profits. The economic rate of return is forward looking and relates the stream of benefits from an investment to the specific asset that generated them. Because of averaging and a backward looking perspective, accounting rates of return cannot equal the eco nomic rate of return. Fisher also argued that even a high correlation between accounting rates of return and the economic rate of return would still create measurement errors that were likely to be correlated with other variables used in industry studies (a correlation which Salamon, 1985, verified). One of the requirements for esti mated coefficients to be unbiased in LINEAR REGRESSION is that the explanatory variables in the regression must not be correlated with the error terms. The systematic measurement error in accounting profits is correlated with variables such as firm size, and therefore studies explaining accounting rates of return by eco nomic variables such as firm size will be mislead ing. In addition, he criticized the use of the rate of return on sales because it fails to properly include economic depreciation, which is a cost of produc tion, and while it may measure market power, it does not measure the economic rate of return.

Edwards, Kay, and Mayer (1987) evaluated several possible adjustments to accounting

procedures that might improve the content of accounting information. They suggested that accountants should report several measures of a firm's profits and asset value since there is no single, unambiguous measure of the "true" values, and different sources of information are required for different purposes. For example, when an asset or activity has a very long life (such as a going concern), the internal rate of return does not necessarily provide useful infor mation because the IRR is a single value defined for the lifetime of an asset and cannot be used to evaluate performance for a fraction of that asset's life. Other measures of profitability from acco unting data may provide investors, economists, and others with better information. Among other suggestions they recommend that value to the owner rules should be used by account ants to value company assets and liabilities where assets are valued (in most cases) at replacement cost. Furthermore, all changes in book values should have entries in the profit and loss state ment and profits and asset values should be adjusted for INFLATION.

Bibliography

Bain, J. S. (1941). The profit rate as a measure of monop-oly power. *Quarterly Journal of Economics*, **55**, 271 93.

Edwards, J., Kay, J., and Mayer, C. (1987). *The Economic Analysis of Accounting Profitability*. Oxford: Oxford University Press.

Fisher, F. M. (1984). The misuse of accounting rates of return: Reply. *American Economic Review*, **74**, 509 17.

Fisher, F. M. and McGowan, J. J. (1983). On the misuse of accounting rates of return to infer monopoly profits. *American Economic Review*, **73**, 82 97.

Long, W. F. and Ravenscraft, D. J. (1984). The misuse of accounting rates of return: Comment. *American Economic Review*, **74**, 494 500.

Martin, S. (1984). The misuse of accounting rates of return: Comment. *American Economic Review*, **74**, 501 6.

Salamon, G. L. (1985). Accounting rates of return. *American Economic Review*, **75**, 495 504.

economics of the Internet

Robert E. McAuliffe

There are several features of the Internet that are unique and give rise to new competition in business. As Shapiro and Varian (1999) point out, the cost of producing information is quite high but the cost of disseminating information is generally quite low. If we combine this with the property of information, you don't know the value of what you don't know, there can be very costly mistakes made with information pricing. Added to that are the SWITCHING COSTS and network effects (*see* NETWORK EX TERNALITIES) that can (or cannot) change demand.

Above all, the Internet lowers the costs of transactions (*see* TRANSACTIONS COSTS). Con sumers can quickly determine a firm's price and how it compares with other Internet or non Internet sellers. By offering such a great amount of information, the Internet changes the pace and quality of business. One overriding concern has been the prospect of digital rights manage ment in such a scenario.

Since information is costly to produce, those who produce that information will be unwilling to give it away, yet they may find it quite difficult to price information appropriately, particularly when that information may be provided pub licly. One solution Shapiro and Varian suggest is to provide *some* of the information free to users so that they can determine whether or not that information would be valuable to them. If it is, the users can pay for all of the information while uninterested users can move on.

The switching costs element of the decision arises because the marginal costs of the infor mation are usually zero, and if producers of information charge a zero price, they will go out of business. If the producers charge a non zero price for information, it is possible that one or more providers of information could charge a price below another's and steal their customers. This creates a dynamic pricing problem for the sellers of information: they must charge a price high enough to cover their costs but not so high that they can lose their customers. The problem is the opposite for the consumer: the consumer wants to pay as little as possible for the infor mation while ensuring that the information will be available in the future.

The last effect that complicates this issue is that of network effects. Network effects occur when it is cheaper or better for a user to buy a product given that other users have already

purchased it. For example, it is not beneficial to be the only owner of a phone; telephone users benefit because almost everyone has a phone. In a similar vein, when one company buys the Windows operating system, it does so in part because so many other companies have already purchased Windows. Again, this creates prob lems for pricing: too low a price and the com pany will not profit, too high a price and the operating system will not sell. These are the problems when pricing products in a networked economy.

Bibliography

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Leibowitz, S. J. (2002). *Rethinking the Network Economy*. New York: American Management Association.

Leibowitz, S. J. and Margolis, S. E. (1995). Path depend ence, lock-in, and history. *Journal of Law, Economics and Organization*, **11**, 205 26.

Shapiro, C. and Varian, H. R. (1999). *Information Rules*. Boston: Harvard Business School Press.

economies of scale

Robert E. McAuliffe

Economies of scale exist when increasing all inputs in production causes output to rise by more than the change in inputs. Therefore if all inputs were doubled (that is, the scale of the firm's operations doubled), output would more than double. Typically, firms vary in terms of *size* and not scale. That is, large firms do not use all inputs in the same proportion as smaller firms. Expressed in terms of costs for a single product firm, the average total costs of produc tion decline as output increases when there are economies of scale (*see* AVERAGE TOTAL COST). These economies of scale may occur at the plant level or the firm level of operations. A firm may operate several plants and there may be economies in the management of the firm or in the consolidation of its financial activities. Economies of scale and ECONOMIES OF SCOPE combined with the level of demand in the market have an important effect on the feas ible number of firms in an industry (Panzar,

1989; *see* MARKET STRUCTURE; MINIMUM EFFICIENT SCALE).

The sources of economies of scale at the firm level are not well understood, although Sharkey (1982) offers several possibilities. At the level of the plant, economies of scale reflect the techno logical or engineering aspects of production. For example, if there are fixed setup costs to produce a specific model on an assembly line, then the per unit setup costs will be lower the greater the number of units produced. Furthermore, given the law of large numbers, when unit sales double, firms do not have to double their inven tories to achieve a given probability of having supplies available to consumers. Firms can also achieve gains from specialization from their labor and capital inputs at higher output levels. In addition, there may be PECUNIARY ECON OMIES of scale when a firm can purchase inputs at lower cost with volume discounts.

For a single product firm, there are econ omies of scale when average cost per unit is declining. When average costs are declining, then marginal costs must be less than average costs (*see* MARGINAL COST). Therefore one measure of economies of scale is the ratio of average costs to marginal costs:

$$s = \text{AC/MC} = \frac{\text{TC}}{Q} \times \frac{\Delta Q}{\Delta \text{TC}} \qquad (1)$$

where TC is the total cost of production, Q is the quantity of output produced, ΔTC is the change in total costs, and ΔQ is the change in output produced. There are economies of scale when $s > 1$ and DISECONOMIES OF SCALE when $s < 1$. Another measure of economies of scale is the cost ELASTICITY, which measures the per centage change in TOTAL COSTS for a given percentage change in output. It can be expressed as:

$$\varepsilon_{\text{TC}} = \frac{\frac{\Delta \text{TC}}{\text{TC}}}{\frac{\Delta Q}{Q}} = \frac{\Delta \text{TC}}{\Delta Q} \times \frac{Q}{\text{TC}} \qquad (2)$$

which is simply the inverse of the scale economy measure in (1) above. There are economies of scale when $\varepsilon_{\text{TC}} < 1$ since costs will increase by less than the increase in output and diseco nomies of scale exist when $\varepsilon_{\text{TC}} > 1$.

A classic empirical study of economies of scale is due to Nerlove (1963), who examined econ omies in electricity generation. He found that economies of scale existed in the industry but were exhausted by the largest firms. Sharkey (1982) and Panzar (1989) provide surveys of both theoretical and empirical issues regarding economies of scale for single product and multi product firms. The measurement and estimation of economies of scale for multiproduct firms is much more complicated than for single product firms because firms may vary their output mix among the different products they produce and therefore will have different costs depending on their output mix (see Baumol, Panzar, and Willig, 1982; Panzar, 1989).

One measure of average costs for a multipro duct firm is *ray average cost* where the compos ition of output (the product mix) is kept constant while the scale of the composite output changes along that ray (see Bailey and Friedlaender, 1982). If a firm produced 100 units of good 1 and 200 units of good 2, ray average costs could then be calculated for any scale of output as long as the proportions of output pro duced remained at 1:2 (e.g., 200 units of good 1 to 400 units of good 2). There are economies of scale for a multiproduct firm when ray average costs are declining as the scale of output increases.

Product specific economies of scale for a mul tiproduct firm can also be measured and occur when the average incremental cost of producing one product decreases, holding the output of all the other products constant. If a firm produces two products (q_1 and q_2), the average incremen tal cost (AIC) is defined as the difference be tween the increase in total costs from producing a given product (q_1) at a specific level of output and the costs incurred if q_1 is not produced at all divided by the total output of q_1. Defining total costs of producing both q_1 and q_2 as $C(q_1, q_2)$, this is calculated as:

$$AIC_1 = \frac{C(q_1,\ q_2) - C(0,\ q_2)}{q_1} \qquad (3)$$

and product specific economies of scale are then measured as the ratio of the average incremental costs for q_1 to the marginal cost of producing q_1:

$$s_1 = AIC_1 / MC_1 \qquad (4)$$

where MC_1 is the marginal cost of producing good 1. Ray scale economies can also be meas ured as the elasticity of output with respect to costs holding the proportion of the two products constant and represent scale economies for a multiproduct firm. Bailey and Friedlaender (1982) suggest that mergers in the trucking indu stry may be motivated by economies of scale and scope once the multiproduct nature of trucking services is recognized.

Bibliography

Bailey, E. E. and Friedlaender, A. F. (1982). Market structure and multiproduct industries. *Journal of Eco nomic Literature*, **20**, 1024 48.

Baumol, W. J., Panzar, J. C., and Willig, R. D. (1982). *Contestable Markets and the Theory of Industry Struc ture*. San Diego: Harcourt Brace Jovanovich.

Martin, S. (2001). *Advanced Industrial Economics*, 2nd edn. Oxford: Blackwell.

Nerlove, M. (1963). Returns to scale in electricity supply. In C. Christ (ed.), Measurement in Economics: Studies in Mathematical Economics and Econometrics. Stan ford, CA: Stanford University Press.

Panzar, J. C. (1989). Technological determinants of firm and industry structure. In R. Schmalensee and R. D. Willig (eds.), *Handbook of Industrial Organization*. New York: North-Holland.

Sharkey, W. W. (1982). *The Theory of Natural Monopoly*. Cambridge: Cambridge University Press.

economies of scope

Robert E. McAuliffe

Economies of scope occur when two or more products can be produced together at lower cost than by producing them separately. Rather than increasing the *scale* of the firm's operations (*see* ECONOMIES OF SCALE), costs are changing with changes in the *scope* of the firm's operations (the number of products produced). Economies of scope may arise when two or more products share a common "public good" input in joint production, where once that common input is acquired it is available to produce other products at no cost (see Bailey and Friedlaender, 1982; Panzar, 1989). There may be economies from shared inputs which give rise to economies of scope when costs are reduced in acquiring those inputs in volume or in sharing them (such as

sharing overhead, reputation, or management). Often knowledge is an important input in production that, once acquired, makes it less costly to produce or sell closely related products. Economies of scale and scope may therefore interact to create advantages for large firms and increase the MINIMUM EFFICIENT SCALE of operation.

If a firm produces two products (q_1 and q_2), the degree of economies of scope can be measured as the difference between the TOTAL COSTS of producing both products separately and the total costs of producing the two products jointly, divided by the total costs of producing the two products jointly. Defining the total costs of producing both q_1 and q_2 jointly as $C(q_1, q_2)$, this is calculated as:

$$S_c = \frac{C(q_1,\, 0) + C(0,\, q_2) - C(q_1,\, q_2)}{C(q_1,\, q_2)}$$

and economies of scope exist when S_c is greater than zero (see Bailey and Friedlaender, 1982; Baumol, Panzar, and Willig, 1982).

Friedlaender, Winston, and Wang (1983) found there were both economies of scale and economies of scope in US automobile manufacturing for large cars, small cars, and trucks. US producers could share common parts and designs between these products and could exploit PECUNIARY ECONOMIES of scale through volume purchases of common parts and materials from suppliers. When the US Department of Justice sought to break up the regional Bell companies from AT&T long distance operations, AT&T initially suggested there were economies of scope in long distance and local phone service and that economic efficiency would be reduced if the NATURAL MONOPOLY in communications were ended. Using Bell system data, Evans and Heckman (1984) tested to determine whether one firm could produce communications output in the US at lower cost than two firms over the period 1958–77. They found that for qualifying data points, costs would be lower if *two* firms produced that output. In other words, the Bell system did not satisfy the necessary cost requirements to be classified as a natural monopoly.

Bibliography

Bailey, E. E. and Friedlaender, A. F. (1982). Market structure and multiproduct industries. *Journal of Economic Literature*, **20**, 1024 48.

Baumol, W. J., Panzar, J. C., and Willig, R. D. (1982). *Contestable Markets and the Theory of Industry Structure*. San Diego: Harcourt Brace Jovanovich.

Evans, D. S. and Heckman, J. J. (1984). A test for sub-additivity of the cost function with application to the Bell system. *American Economic Review*, **74**, 615 23.

Friedlaender, A. F., Winston, C., and Wang, K. (1983). Costs, technology and productivity in the US automobile industry. *Bell Journal of Economics*, **14**, 1 20.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Panzar, J. C. (1989). Technological determinants of firm and industry structure. In R. Schmalensee and R. D. Willig (eds.), *Handbook of Industrial Organization*. New York: North-Holland.

Sharkey, W. W. (1982). *The Theory of Natural Monopoly*. Cambridge: Cambridge University Press.

efficiency

Robert E. McAuliffe

Economists use the criterion of Pareto optimality to determine whether production or consumption decisions are efficient (*see* PARETO OPTIMAL ALLOCATION). According to this criterion, an allocation is efficient if in a society no person can be made better off without making someone else worse off. This concept of efficiency treats consumer tastes and the distribution of income as given, and efficiency is achieved when society produces those goods which consumers are willing to pay for at the lowest possible cost. Given the general principle of Pareto optimality, several specific measures of efficiency are provided below for an economy that produces two goods, X and Y, using two inputs in production, labor and capital (L and K, respectively) for two consumers, A and B (see Pindyck and Rubinfeld, 1995; Baumol, 1977). These efficiency conditions can be generalized for economies with many products, consumers, and resources (see Baumol, 1977).

ALLOCATIVE EFFICIENCY

Allocative efficiency exists when those consumers who are willing and able to pay the marginal cost to society of producing a product receive it. This requirement is fulfilled when the market price (P) equals MARGINAL COST (MC)

as long as there are no EXTERNALITIES in pro duction or consumption. When a consumer maximizes TOTAL UTILITY, the MARGINAL RATE OF SUBSTITUTION of good Y for good X (MRS_{XY}) is equal to the price ratio, P_X/P_Y. If markets are perfectly competitive (*see* PERFECT COMPETITION), then all consumers face the same price ratio and if they also maximize their utility, then the marginal rates of substitu tion of good Y for good X will be equal across all consumers of the two goods. Thus efficiency in exchange requires:

$$MRS_{XY}^{A} = MRS_{XY}^{B}$$

where MRS_{XY}^{j} is the marginal rate of substitu tion of good X for good Y for consumer j. If this condition does not hold, then the consumers could both benefit from exchanging the goods until their marginal rates of substitution were equal.

PRODUCTIVE EFFICIENCY

Productive efficiency requires resources in pro duction to be allocated so that output in the economy is maximized at the lowest cost. As Baumol (1977) has suggested, productive effi ciency is simply the application of Pareto optim ality to production, where for any arbitrarily selected product, output of that product is maxi mized subject to the constraint that there is no reduction in the output of all the other goods produced. Productive efficiency is necessary for Pareto optimality but it is not sufficient since society could very efficiently produce products which consumers did not want.

The resources used to produce output will be efficiently employed across different industries in the economy when the MARGINAL RATE OF TECHNICAL SUBSTITUTION of labor for cap ital ($MRTS_{LK}$) in one industry (X) equals its value in other industries (Y). For the two good, two resource economy in this example, the con dition for productive efficiency is:

$$MTRS_{LK}^{X} = MTRS_{LK}^{Y}$$

Here resources will be efficiently employed be tween the two industries when they are equally productive in both. Again, if the markets are perfectly competitive, then firms in both indus tries will maximize profits by hiring labor and capital until the marginal rate of technical sub stitution of labor for capital equals the ratio of the labor wage rate to the cost ("rental" rate) for capital. Since all industries face the same wage and capital rental costs in a perfectly competitive economy, the marginal rates of technical substi tution will be equal across firms in all industries.

Firms in competitive industries maximize profits by producing until price (P) equals mar ginal cost (MC), and this coordinates productive efficiency with allocative efficiency because the price measures the value of a unit of the product to consumers in society, assuming there are no externalities in consumption, while marginal cost reflects the opportunity cost to society of producing that product, if there are no external ities in production (*see* OPPORTUNITY COSTS). Therefore competitively determined prices coor dinate the production and consumption deci sions in society so that those consumers who value products the most and are willing and able to pay the costs to society of producing them will receive them. When these efficiency conditions hold, resources cannot be reallocated to make anyone better off without making some one else worse off.

See also *Pareto optimal allocation; profit maxi mization; utility maximization*

Bibliography

Bator, F. M. (1957). The simple analytics of welfare maximization. *American Economic Review*, **47**, 22 59.
Baumol, W. J. (1977). *Economic Theory and Operations Analysis*. Englewood Cliffs, NJ: Prentice-Hall.
Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.
Pindyck, R. S. and Rubinfeld, D. L. (1995). *Microecono mics*, 3rd edn. Englewood Cliffs, NJ: Prentice-Hall.

efficient markets hypothesis

Vickie L. Bajtelsmit

In the most general sense, an efficient financial market is one that allocates funds to their most productive uses and, through COMPETITION, results in market prices that are consistent with the underlying value given the information

available at the time (*see* EFFICIENCY). Thus, efficient capital markets are those in which the market price of each security is a good indicator of the firm's future prospects.

In a perfect market, the value of any security will reflect all information that is available to the participants in the marketplace. Therefore, cap ital market efficiency implies that:

1 securities are always fairly priced;
2 their prices react quickly to new information; and
3 investors cannot make excess returns simply by using strategies based on available infor mation.

As an example, suppose that a firm announces that it has been successful in a bid to buy a company that supplies one of the component parts used in its production process and, as a result, management anticipates increased earn ings in the future. If investors believe this an nouncement, their estimates of the value of the company should rise and the stock price should rise. In an inefficient market, an investor would be able to act on this information to gain excess profits by buying the stock immediately upon hearing the announcement and then selling it when the price has increased as a result of the announcement. In a perfectly efficient market, by the time the investor can call his or her broker, the stock price will already have risen to reflect the new information. Clearly, however, in order for the price to adjust, it must be the case that at least one investor is able to capitalize on the new information. The more efficient the market, therefore, the quicker the information is impounded in prices and the lower the oppor tunity to capitalize on new information.

Although the theory of market efficiency as applied to the capital markets has its origins in basic economic theory, Eugene Fama (1970) is often credited as the first to distinguish several levels of efficiency by the *type* of information that is reflected in prices. His analysis of effi ciency on this basis has come to be known as the "efficient markets hypothesis" and is subdivided into three levels of efficiency. In the *weak form* of efficiency, prices of securities fully reflect all information contained in historical price move ments. In the *semi strong form* of efficiency,

security prices are said to reflect all publicly available information. Finally, in a *strong form* efficient capital market, prices fully reflect all information, both public and private. Tests of market efficiency have generally attempted to determine the ability of investors to make excess profits using the sets of information identified by the theory: price data only; all publicly available data; or all information.

Some investors believe that new information is incorporated in stock prices slowly and that they can earn an excess return by trading based on careful analysis of generally available infor mation. Investors that use "technical analysis," the observation of patterns in past price move ments in an attempt to predict future price movements, believe that the market is not even weak form efficient. If seasonal or annual cycles in prices exist or if price changes are correlated over time, an investor might be able to devise a trading rule that produces consistent returns (by buying at the low end of the price cycle and selling at the high end). However, for this to be an effective strategy (and thus proof of market inefficiency), excess returns achieved by the trading rule would have to be sufficient to cover trading costs. Tests of the weak form of efficiency have generally shown that trading rules based on perceived patterns in historical returns do not earn excess returns for investors that are sufficient to cover trading costs.

Although information about past price move ments may be instantaneously incorporated in prices, it does not necessarily follow that other information is as well. Therefore, tests of the semi strong form of efficiency have focused on a variety of types of publicly available informa tion. The methodology, initially introduced by Fama, Fisher, Jensen, and Roll (1969) in testing for the effects of stock splits and dividend announcements on security prices, has been applied to many different types of information. Although the techniques have been refined over the years, the conclusions have generally been similar. The market seems to adjust to new in formation rapidly and much of the impact on price takes place in anticipation of the actual public announcement. Although this supports semi strong efficiency, it is still possible for in vestors to make excess profits if they can acquire information that is not widely known and act on

it before the information becomes incorporated in prices.

Some investors, such as managers or other insiders, clearly would have an advantage in capitalizing on non public information. In fact, the only way that all information, both public and private, could be incorporated in market prices (i.e., semi strong efficiency) would be if insiders were using their private information in trading. In the example above, if the managers purchased shares in their own company prior to making the public announcement and then sold the shares for a gain after the announcement caused stock prices to rise, they would make an excess profit. Insider trading laws make this type of activity illegal, although there is evidence that such fraudulent trading has occurred without detection by the authorities.

Bibliography

Elton, E. J. and Gruber, M. J. (1991). *Security Analysis and Portfolio Theory*, 4th edn. New York: John Wiley.

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, **25** (2), 383 417.

Fama, E., Fisher, L., Jensen, M., and Roll, R. (1969). The adjustment of stock prices to new information. *International Economic Review*, **10** (1), 1 21.

elasticity

Gilbert Becker

This measure indicates the responsiveness of a buyer or seller to a change in the value of an economic variable. Several types of elasticity exist, although perhaps the most important type is *price elasticity of demand*. This measure indicates the extent to which consumers of a product react, in terms of the quantity of the good which they purchase, when there is a change in the price of that product. Additional measures including price elasticity of supply, CROSS ELASTICITIES, INCOME ELASTICITY, and CROSS ADVERTISING ELASTICITY are also important in determining the STRATEGIC BEHAVIOR of firms and outcomes in markets. Price elasticity of demand (ε_p) is calculated as the percentage change in quantity demanded resulting from a small percentage change in price. Written in the form of a fraction this becomes:

$$\varepsilon_p = \frac{\Delta Q}{Q} \bigg/ \frac{\Delta P}{P} = \frac{\Delta Q}{\Delta P} \times \frac{P}{Q}$$

where the symbol Δ indicates a change in quantity (Q) or price (P). Where the exact DEMAND FUNCTION is already known, an infinitesimally small change in price can be used to calculate what is known as the point elasticity. More commonly, when a manager makes discrete changes in price, the estimation of elasticity involves a range of prices and quantities. Calculating elasticity using the average price and quantity over these ranges yields a useful approximation (*see* ARC ELASTICITY).

As the LAW OF DEMAND indicates that price and quantity demanded are inversely related, the value of ε_p should always be negative. Since the negative sign adds no information about the size of the reaction, by convention it is typically dropped. Three levels of responsiveness are observable. When the percentage change in quantity demanded exceeds the percentage change in price, the absolute value of ε_p has a value greater than 1 and demand is said to be *elastic* with respect to price. When $|\varepsilon_p| < 1$, a small percentage change in quantity has occurred, which is the case of *inelastic* demand, while the situation where the percentage change in price and quantity exactly offset each other, $|\varepsilon_p| = 1$, is identified as *unit elastic* demand.

The degree of consumer sensitivity to a price change is influenced by several factors. Perhaps most apparent is that the greater the extent to which a product is perceived as being a necessity, the more inelastic its demand tends to be with respect to price. A rise in the price of a product will induce only a small reduction in sales if consumers feel that the good is truly needed.

In addition, switching to alternative products is common when one product's price is changed. Demand for a product thus tends to become more elastic where the number and closeness of available SUBSTITUTES rise, as this substitution of products more readily and easily occurs. Nelson (1974) in his study of ADVERTISING cautions that elasticity actually depends on the number of alternatives of which the consumer is aware. Managers need to consider this and the

appropriate market boundaries (*see* MARKET DEFINITION) in determining their business strategy. In PERFECT COMPETITION, where large numbers of rivals produce identical prod ucts, demand for any one firm's product is per fectly elastic and all firms become price takers. Where a strategy of PRODUCT DIFFERENTI ATION is possible, it can be used to create a more inelastic demand for a product. One well known example is that of 7UP being marketed as the "UNCOLA" in an effort to distinguish itself from its rivals.

Similarly, a product requiring a greater por tion of the consumer's budget will tend to have a more elastic demand as a result of the stronger INCOME EFFECT from a price change. A small percentage decline in price may result in a large increase in sales of high expense items such as automobiles or major household appliances due to the significant overall savings felt by con sumers. This increase in sales is reinforced by the fact that the rewards from search are higher when the good in question takes up a significant percentage of consumer income (*see* SEARCH COSTS). Finally, demand tends to become more elastic over longer time periods. This results from the fact that, typically, a wider array of substitutes becomes available over time. Also, a consumer simply has more time to react to the price change.

Price elasticity of demand is central to a firm's sales revenues. Since a firm's TOTAL REVENUE is dependent on the product's price and the number of units sold, the impact of a price change on revenues depends singularly on the product's elasticity of demand. As the law of demand requires that price and quantity demanded move in opposite directions, a change in price creates two opposing forces on revenues. For example, a lower price brings in additional revenue from the additional units now being sold, but at the same time brings in less revenue on the units which would have been sold at the previous price. The elasticity of demand indicates which force is stronger. A price cut in the face of an elastic demand for the product will increase the total receipts from its sales because of the large and more than offsetting rise in unit sales. Similarly, a price rise where demand is elastic will reduce total revenues.

Price elasticity of demand also plays a vital role in determining optimal prices. For PROFIT MAXIMIZATION, a firm should produce until the MARGINAL REVENUE (MR) equals the MARGINAL COST (MC). For linear demand curves it can be shown that marginal revenue is related to price (P) and elasticity by the equation

$$MR = P \times [1 + (1/\varepsilon_p)]$$

As such, if a firm maximizes profits, marginal revenue in the equation above equals marginal cost, and this relation can be used to show that the best price to charge is dependent on both elasticity and cost. Maurice and Thomas (2002) show that using a cost plus MARKUP PRICING strategy having the general form of

$$P = MC \times (1 + markup)$$

can yield, under certain circumstances,

$$P = MC + [-1/(\varepsilon_p + 1)] \times MC$$

in which the profit maximizing markup of price is a percentage of cost and is inversely related to the product's demand elasticity. That is, the more elastic is the demand for the product, the lower will be its profit maximizing markup price.

See also *Lerner index; market power*

Bibliography

Browning, E. and Zupan, M. (2004). *Microeconomics: Theory and Applications*, 8th edn. New York: John Wiley.

Maurice, S. and Thomas, C. (2002). *Managerial Econom ics*, 7th edn. Boston: McGraw-Hill Irwin.

Miller, R. L. (2004). *Economics Today: The Micro View*, 12th edn. Boston: Pearson.

Nelson, P. (1974). Advertising as information. *Journal of Political Economy*, **82**, 729 54.

entry

Robert E. McAuliffe

Entry occurs when new firms engage in the production of specific goods or services in an industry and represents an important feature of competition in market economies. New firms

may enter an industry either as wholly new enter prises or as extensions of existing firms from other markets. Entry is the crucial mechanism which disciplines existing firms within an indus try. In the extreme case where there is a perfectly contestable market, even a monopolist cannot earn economic profits or incur excessive costs because it will be subjected to hit and run entry that will force the market price down until it equals average cost (Baumol, Panzar, and Willig, 1982; *see* CONTESTABLE MARKETS; ECO NOMIC PROFIT). However, potential entrants may be prevented from entering the industry when there are SUNK COSTS or other BAR RIERS TO ENTRY. Weitzman (1983) argued that within short periods of time some costs are always sunk, and this has led some to question how much discipline can be imposed on estab lished firms by potential entry (see also Shep herd, 1984). Barriers to entry may prevent or reduce entry into a market even when economic profits exist. The sources of barriers to entry may also interact and have significant joint effects in deterring entrants, although individu ally each source may not be substantial (see, e.g., Geroski, Gilbert, and Jacquemin, 1990).

Economic theory predicts entry should occur in response to expected economic profits in a given market which are determined by the en trant's costs, the expected response of existing firms to the entrant, and the post entry price and market share the entrant expects. Porter (1980) discusses several considerations regarding the entry decision, including the anticipated reac tion of established firms to the entrant. Among other issues he mentions several factors that increase the likelihood that existing firms will retaliate, such as slow industry growth, high FIXED COSTS, high industry concentration, and products which are not differentiated (*see* PRODUCT DIFFERENTIATION). When indus try concentration is high or when industry sales are slow, an entrant is more likely to take sales from existing firms and trigger a response. High fixed costs and products which are not differen tiated raise the probability of a price war because all firms have greater incentives to cut prices. With high fixed costs any sale will make *some* contribution to overhead, so firms are motivated to reduce their price to make an additional sale, whereas products which are not differentiated

have many SUBSTITUTES and consumers are more likely to make purchases based on price.

Entry into an industry may also be motivated by anticipated market growth and technological change in addition to expected economic profits. New products introduced by innovative firms are part of the dynamic process of COMPETI TION that Schumpeter (1975) called the "gale of creative destruction." Important as it is, the analysis of entry rates is complicated by the fact that existing firms may pursue strategies such as LIMIT PRICING to prevent or reduce the rate of entry. In their survey, Geroski et al. (1990) examined the empirical literature on entry in terms of how quickly it responded to industry profits and how rapidly it reduced those profits. They found that entry rates were low, suggesting high barriers to entry, and that even when entry was rapid, industry profits remained high. As they observed, conclusions regarding the rela tionship between entry and profits are subject to several qualifications. One problem is that ac counting profits do not necessarily reflect eco nomic profits, even when the two are highly correlated (Fisher and McGowan, 1983; *see* ACC OUNTING PROFIT). Another point they raise is that the threat of entry may force existing firms to eliminate X inefficiency and reduce costs in production (*see* X EFFICIENCY). If existing firms reacted this way, then industry profits could remain constant despite the fact that entry had significant effects on firm behav ior.

In a careful study of entry and EXIT in US manufacturing over the period from 1963 to 1982, Dunne, Roberts, and Samuelson (1988) distinguished three types of entry. Entry into an industry could result from a new entrant constructing a new plant, a diversifying firm from another industry constructing a new plant, or an existing firm changing its product mix. They found that, on average, 38.6 percent of firms in operation in each industry had not produced in that industry in the previous census. Furthermore, new entrants with new plants tended to be small relative to the size of existing firms and accounted for a significant portion of the number of entrants into an indus try, but their average market share was small. Diversifying firms tended to be larger and were more likely to survive. While single plant firms

accounted for 93.4 percent of the total number of firms in each year, their share of the value of production was only 17.1 percent. On average, most of the value of industry production was produced by multiplant firms. One interesting discovery by Dunne et al. was that industries with high rates of entry also tended to have high rates of exit. The correlation between the rate of entry into an industry and exit from the industry was very high in terms of both the numbers of entrants and exiters and their market share. In addition, the differences in entry and exit patterns between industries persisted over time, which suggests that industry specific factors affect both entry and exit.

Bibliography

Baumol, W. J., Panzar, J. C., and Willig, R. D. (1982). *Contestable Markets and the Theory of Industry Struc ture*. San Diego: Harcourt Brace Jovanovich.

Dunne, T., Roberts, M. J., and Samuelson, L. (1988). Patterns of entry and exit in US manufacturing indus tries. *The Rand Journal of Economics*, **19**, 495 515.

Fisher, F. M. and McGowan, J. J. (1983). On the misuse of accounting rates of return to infer monopoly profits. *American Economic Review*, **73**, 82 97.

Geroski, P., Gilbert, R. J., and Jacquemin, A. (1990). *Barriers to Entry and Strategic Competition*. New York: Harwood.

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.

Schumpeter, J. A. (1975). *Capitalism, Socialism, and Dem ocracy*. New York: Harper and Row.

Shepherd, W. G. (1984). "Contestability" vs. competi tion. *American Economic Review*, **74**, 572 87.

Weitzman, M. L. (1983). Contestable markets: An upris ing in the theory of industry structure: Comment. *American Economic Review*, **73**, 486 7.

equilibrium

Robert E. McAuliffe

The concept of equilibrium is fundamental to economic analysis. As with physical systems, an economic system is in equilibrium when it is at rest and exhibits no tendency to change until the underlying market conditions change. Markets are said to be in equilibrium when the quantity supplied equals the quantity demanded and this determines the equilibrium price and quantity in the market. Once equilibrium is achieved in a market, consumers are able to purchase the amount they wish to buy at the equilibrium price and firms are able to sell the amount they wish to sell. Since both consumers and firms are able to do what they wish, given the actions of others in the market, they have no reason to change their behavior and so the equilibrium is maintained until cost or demand conditions change. Given this, Hayek (1948: 42) suggested that the concept of equilibrium should be extended to require that the plans of economic agents are fulfilled given the behavior of others in the market. High (1990) elaborated Hayek's definition and argued that equilibrium occurs when the various plans made by economic agents are mutually compatible (perhaps even inter temporally).

Closely related to the concept of equilibrium is the stability of equilibrium. A stable equilib rium is one to which the economic system will return if, for any reason, it is slightly disturbed from that equilibrium. Fisher (1983) argued that the conclusions reached with equilibrium analy sis are only reliable to the extent that it can be shown that markets converge to new, stable equilibrium positions reasonably quickly. How ever, if consumers and firms trade out of equi librium, convergence to the equilibrium price may not be achieved. He showed that, with some significant restrictions, markets could con verge to equilibrium over time when agents trade out of equilibrium.

Another closely related issue is the uniqueness of equilibrium. If demand equals supply at one price and quantity combination, it may be pos sible for the same supply and DEMAND CURVES to have other equilibria with no changes under lying demand or supply. Although a non unique equilibrium is always a possibility, most supply and demand curves estimated in practice are linear or log linear and this rules out the option of multiple equilibria.

Critics of equilibrium analysis in economics have argued that there is no convincing theory describing how markets could achieve equilib rium and that markets may never reach it (see Nelson and Winter, 1982; Blaug, 1992; Haus man, 1992). But as High has noted, even if markets never reach equilibrium, it still remains

a useful theoretical concept to predict how markets might respond to specific changes. For example, an increase in demand will cause a complex series of adjustments in the market which are not well understood. By focusing on the new equilibrium, economists abstract from the dynamic process of market adjustment al together to determine where the market process may lead if there are no other changes. High also argues that even if markets do not reach equilibrium, that does not mean that there are no equilibrating forces in the market.

As an alternative to equilibrium analysis, Nelson and Winter (1982) recommend evolu tionary modeling where agents in a market develop strategies, react to their environment, learn, and adapt over time. They argue that firms and consumers do not necessarily maxi mize in new and unfamiliar situations because they have not learned how best to operate in a new environment. Lesourne (1992) developed several simple models where agents learn and adapt to the market environment, and he found that in many cases when agents had IMPERFECT INFORMATION or there were TRANSACTIONS COSTS, convergence to "the" market equilib rium of neoclassical theory did not occur. He found that markets could reach stable equilibria which were not necessarily optimal and the path taken to reach equilibrium could affect the equi librium ultimately attained. In this respect the problems of convergence to equilibrium identi fied by Lesourne were alluded to in Fisher's (1983) discussion of the behavior of markets in disequilibrium.

Although evolutionary modeling offers an appealing alternative to equilibrium analysis, it also creates additional problems. As Blaug (1992) observes, if the assumption of rational, maximizing agents is abandoned, how can in complete information and incorrect expectations be modeled? Since firms and consumers can have an infinite variety of incorrect expectations, a Pandora's box of possibilities is opened once the assumption of maximizing behavior in equi librium is sacrificed.

Several equilibrium concepts which fre quently appear in economic applications are dis cussed below.

1   STATIC VS. DYNAMIC EQUILIBRIUM

A static equilibrium is one which occurs at a single point in time, such as the equality between the quantity supplied and demanded for a good. A dynamic equilibrium is one which occurs over time, such as a predicted increase in the price of a product. If the price increase is correctly antici pated by agents in the market, then no one is surprised as the price increases each period and the market is in a dynamic, though possibly changing, equilibrium from one period to the next.

2   SHORT RUN VS. LONG RUN EQUILIBRIUM

An industry is in short run equilibrium when short run supply (holding the number of firms in the industry fixed) equals demand. In the short run the number of firms is fixed, so there may be economic profits or losses during this period (*see* ECONOMIC PROFIT). However, in the long run firms may enter or exit the industry, so long run equilibrium occurs when there is no net entry into or exit from the industry. For a perfectly competitive industry, the entry and exit of firms will cause economic profits to be zero in the long run (*see* CONTESTABLE MARKETS; ENTRY; EXIT; PERFECT COMPETITION).

3   NASH EQUILIBRIUM

Frequently employed in GAME THEORY, a Nash equilibrium is one where no agent wants to change his or her behavior given the behavior of others in the market. For firms in an OLI GOPOLY, a Nash equilibrium in strategy occurs when no firms want to change their strategies (ADVERTISING policy, pricing, output, research and development, etc.) given the strategies of the other firms in the industry.

See also *Cournot competition; strategic behavior*

Bibliography

Blaug, M. (1992). *The Methodology of Economics*, 2nd edn. New York: Cambridge University Press.

Fisher, F. M. (1983). *Disequilibrium Foundations of Equi librium Economics*. Cambridge: Cambridge University Press.

Hausman, D. M. (1992). *The Inexact and Separate Science of Economics*. New York: Cambridge University Press.

Hayek, F. A. (1948). *Individualism and Economic Order*. Chicago: University of Chicago Press.

High, J. (1990). *Maximizing, Action, and Market Adjustment*. Munich: Philosophia Verlag.

Lesourne, J. (1992). *The Economics of Order and Disorder*. New York: Oxford University Press.

Nelson, R. R. and Winter, S. G. (1982). *An Evolutionary Theory of Economic Change*. Cambridge, MA: Harvard University Press.

error terms in regression

Robert E. McAuliffe

The estimated coefficients from a LINEAR REGRESSION will be the best linear unbiased estimators (BLUE) when the error terms in the regression have the properties described below. To illustrate, suppose a manager wanted to estimate the relationship between unit sales of a product, Y_i, and advertising, X_i. The regression relationship is assumed to be linear and is written as

$$Y_i = \beta_0 + \beta_1 \times X_i + v_i \qquad (1)$$

where the Y_i are observations on the *dependent variable*, X_i are observations on the *independent* or *explanatory variable*, and the v_i are the error terms. The coefficients β_0 and β_1 are to be estimated and will show how much effect ADVERTISING has on sales. If the coefficient for β_1 is not significantly different from zero, then advertising has no effect on this firm's sales. Additional explanatory variables could easily be added, but to simplify the discussion we will assume only one explanatory variable to explain sales below.

The disturbances in the regression v_i are added to reflect three possible sources of randomness. First is the likely random nature of the relationship between advertising and sales (i.e., the link between these two variables will not be perfect), and so a perfect fit between advertising and sales is unlikely. Second, the effects of other explanatory variables which are missing from the regression are reflected in the error terms. Finally, there may be measurement errors in the observations on the dependent variable which the error term will capture (see Maddala, 2001; Greene, 2002).

For linear regression estimates of the coefficients in (1) above to be BLUE, the error terms must satisfy the following assumptions:

1. The errors must have zero mean, so their EXPECTED VALUE is zero: $E(v_i) = 0$.
2. The errors must have a constant VARIANCE: $var(v_i) = \sigma^2$.
3. There is no AUTOCORRELATION in the error terms. This means that successive observations of v_i are not correlated with each other, so $cov(v_i, v_j) = 0$ for all $i \neq j$.
4. The error terms are independent of all of the explanatory variables. That is, the COVARIANCE between the explanatory variables and the errors is zero: $cov(v_i, X_j) = 0$ for all i and all j. Violation of this assumption will cause the linear regression estimates to be biased.
5. The error terms are independently and normally distributed. Thus, $v_i \sim IN(0, \sigma^2)$.

When the error terms in a linear regression satisfy these assumptions, then the least squares estimators will be unbiased and efficient in the class of all linear unbiased estimators. Furthermore, with assumption 5 statistical hypothesis tests regarding the estimated coefficients can be performed using F tests and T STATISTIC tests in small samples (see Maddala, 2001; Greene, 2002). If the error terms fail to satisfy these assumptions, it may be possible to correct the problem. For example, most regression software allows users to correct for first order serial correlation in the errors and offers procedures to correct for HETEROSKEDASTICITY (the failure of assumption 2).

See also *multicollinearity*

Bibliography

Greene, W. H. (2002). *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.

Gujarati, D. N. (2002). *Basic Econometrics*, 4th edn. New York: McGraw-Hill Irwin.

Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.

estimating demand

Robert E. McAuliffe

When prices and output in a market reflect the forces of both supply and demand, estimates of the demand (or supply) curve by LINEAR RE GRESSION will be biased (*see* SIMULTANEOUS EQUATIONS BIAS). This bias occurs because in a simultaneous equations system, the regressors will be correlated with the error term (*see* ERROR TERMS IN REGRESSION). For example, con sider the system of market supply and demand equations below:

$$Q_t^d = \alpha_1 + \alpha_2 P + \alpha_3 I + \varepsilon_t^d \qquad (1)$$

$$Q_t^s = \beta_1 + \beta_2 P + \beta_3 C + \varepsilon_t^s \qquad (2)$$

$$Q_t^d = Q_t^s \qquad (3)$$

where Q is the quantity supplied or demanded, P is the price of the product, I is the consumers' real income (which is an exogenous variable in this model), C is the resource cost of production (also exogenous), and the ε_t^i are random disturb ances to demand and supply. Equations (1)–(3) are referred to as *structural equations* which re flect specific features of the market. These equa tions are mutually dependent, so random variations in demand or supply will cause random variations in price. But if demand is randomly high (or low) in a given period, the price will be higher (lower) in that period as well, and this means there is a correlation between one of the right hand side variables in the regression (price) and the error term which biases linear regression estimates. The same problem occurs with estimates of the supply curve or with esti mates of any simultaneous relationship where variables on the right hand side of the regres sion are not exogenous. There are several methods for obtaining consistent, unbiased esti mates of simultaneous equations.

INDIRECT LEAST SQUARES

One method of estimating demand or supply in a system of equations such as (1)–(3) is to estimate the *reduced form parameters* – those parameters derived from the structural equations where the dependent variables are expressed in terms of the exogenous variables only. For the three equation system above, price and quantity can be expressed solely in terms of the exogenous variables:

$$Q = \pi_1 + \pi_2 I + \pi_3 C + \eta_1 \qquad (4)$$

$$P = \pi_4 + \pi_5 I + \pi_6 C + \eta_2 \qquad (5)$$

where the π_i are the reduced form parameters and the η_i are error terms. Since the right hand side variables in equations (4) and (5) are exog enous, a linear regression can be used to esti mate the π_i coefficients. If the equations in the system are exactly identified, there is a unique relationship between each of the structural coef ficients (the αs and βs) and the reduced form parameters (the π_is) which enables the re searcher to solve for the structural parameters (*see* IDENTIFICATION PROBLEM). When an equation is over identified, there may be more than one reduced form estimate for some of the structural coefficients. Thus the parameters of the structural equations are estimated "indir ectly" from the estimates of the reduced form parameters.

For example, in the system (1)–(3) above, solving for Q and P as functions of the exogenous variables only, we have:

$$Q = \frac{\alpha_1\beta_2 - \alpha_2\beta_1}{\beta_2 - \alpha_2} + \frac{\alpha_3\beta_2}{\beta_2 - \alpha_2} \times I - \frac{\alpha_2\beta_3}{\beta_2 - \alpha_2} \times C + \text{error term}$$

$$(6)$$

$$P = \frac{\alpha_1 - \beta_1}{\beta_2 - \alpha_2} + \frac{\alpha_3}{\beta_2 - \alpha_2} \times I - \frac{\beta_3}{\beta_2 - \alpha_2} \times \quad (7)$$
$$C + \text{error term}$$

which means that:

$$\pi_1 = \frac{\alpha_1\beta_2 - \alpha_2\beta_1}{\beta_2 - \alpha_2},$$

$$\pi_2 = \frac{\alpha_3\beta_2}{\beta_2 - \alpha_2}, \qquad (8)$$

$$\pi_3 = \frac{\alpha_2\beta_3}{\beta_2 - \alpha_2},$$

and so on. We can solve for the structural param eters using the reduced form estimates by matching the terms in (6) and (7) with those in (4) and (5). Thus the "indirect" estimates are:

$$\hat{\alpha}_1 = \hat{\pi}_1 - \hat{\alpha}_2\hat{\pi}_4 \text{ and } \hat{\alpha}_2 = \frac{\hat{\pi}_3}{\hat{\pi}_6}, \text{ etc.}$$

The indirect least squares method is not widely employed in research but it is useful in demonstrating the identification problem (for additional information see Maddala, 2001; Greene, 2002).

INSTRUMENTAL VARIABLES AND TWO-STAGE LEAST SQUARES

The instrumental variables method and two stage least squares are "limited information" estimation techniques and are more frequently employed to avoid simultaneous equations bias. They are called limited information methods because each equation in the system is estimated separately. Full information methods estimate the entire system of equations and allow interaction between the equations but are beyond the scope of this entry.

Since the problem in estimating demand or supply is the correlation between one of the regressors (price) and the error term, one solution is to replace the endogenous variable (price) with a substitute variable which is highly correlated with it but is not correlated with the error term. The demand or supply equations are then estimated using the substitute variable. This is the essence of the instrumental variables and two stage least squares techniques.

Given the system of equations (1)–(3), the income and resource cost variables, I and C, are exogenous (uncorrelated with the error term) and should be highly correlated with price because the market price will change whenever these variables change. So I and C are both candidates to serve as instruments. Thus to construct a substitute variable for price in the demand curve, the researcher should regress price on the exogenous variable, C. The resulting fitted values for price, $\hat{P}$, from that regression will then replace price in the original demand curve. These fitted values will be highly correlated with the original price variable but they will also be uncorrelated with the error term in the demand equation by construction. A similar procedure can be used to estimate the supply curve, equation (2) above, using income as the instrument. The demand or supply curve can then be estimated by regressing quantity on the fitted price variable and the exogenous variables. Thus the researcher estimates:

$$Q_t^d = \alpha_1 + \alpha_2\hat{P} + \alpha_3 I + \varepsilon_t^d \qquad (9)$$

$$Q_t^s = \beta_1 + \beta_2\hat{P} + \beta_3 C + \varepsilon_t^s \qquad (10)$$

Although this regression with fitted values is not the same as the original structural model (1) and (2), the structural parameters will be consistently estimated in large samples using this technique. By contrast, estimates from a linear regression will be biased no matter how large the sample (see Greene, 2002). A number of variables could be used as instruments when an equation is over identified and the instrumental variable method offers little guidance in such cases.

The two stage least squares technique also uses instrumental variables for exactly and over identified equations. However, with two stage least squares *all* of the exogenous and predetermined variables in the model are used as instruments in the first stage of the regression. The structural parameters are then estimated in the second stage regression. This procedure is available in most statistical and econometric software programs and weights each instrument optimally. This can be important when estimating over identified equations where several instrumental variables could be used.

Two studies which employed the instrumental variables and two stage least squares methods to estimate the demand for cigarettes in the US were Bishop and Yoo (1985) and McAuliffe (1988). In contrast with previous studies which used biased linear regression methods, these authors found that the demand for cigarettes was inelastic with respect to price (*see* ELASTICITY) and that the 1964 Surgeon General's report on the negative health effects from smoking significantly lowered the demand for cigarettes. However, the 1971 ban on broadcast advertising of cigarettes on radio and television in the US had no independent statistical effect.

Bibliography

Bishop, J. A. and Yoo, J. H. (1985). Health scare, excise taxes and advertising ban in the cigarette demand and supply. *Southern Economic Journal*, **52**, 402 11.

Greene, W. H. (2002). *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.

Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.

McAuliffe, R. E. (1988). The FTC and the effectiveness of cigarette advertising regulations. *Journal of Public Policy and Marketing*, 7, 49–64.

EU competition policy, 2004

Gilbert Becker

European Union (EU) policy regarding COMPETITION holds as a primary objective the promotion of the economic integration of the region. The law in some respects parallels US antitrust law in that it prohibits artificial restrictions on competition and proscribes abuses of MARKET POWER. It also tends to promote the goals of fairness and the protection of consumer interests with respect to prices and choice. These latter objectives, and that of sustaining small and medium sized firms as an end in and of itself, appear to be more prominent goals than in the US.

Unlike the US, which has had its antitrust laws in place for nearly all of the last century, antitrust law is relatively new for much of the rest of the world. Palim (1998) finds that of the 70 countries having such laws in place by 1996, nearly 80 percent originated after 1980. While antitrust law applying to the EU and its predecessors dates back to the 1957 Treaty of Rome, over half of the original 15 members (e.g., Italy, Ireland, Sweden) only established their own national laws after 1989. Moreover, these laws remain in a state of flux as modernization of national laws (e.g., the UK's 1998 Competition Act) continues, often in an effort to conform to EU law. Finally, the EU is currently meeting in a constitutional convention, and with the accession of ten new member states in 2004, is continuing a process of revisions of the region's current antitrust policy. (Some revisions are complete and will be examined below; *see* EU MERGER GUIDELINES, 2004; EU MERGER POLICY, 2004.)

Recent EU policy on competition and its regulation stems from Articles 81 and 82 of the 1999 Amsterdam Treaty. Article 81 makes "prohibited as incompatible with the common market" agreements which "may affect trade between the member states" and which have the effect of "prevention, restriction or distortion of competition" (see European Communities, 1997, for a complete statement on these laws and their implementation). Several specific business practices are identified, including agreements which:

1 fix prices or other trading conditions;
2 limit production, investment, or research and development;
3 share markets; or
4 place trading partners at a competitive disadvantage by discriminating in conditions on otherwise equivalent transactions.

This law parallels Section 1 of the US SHERMAN ACT, which proscribes combinations "in restraint of trade," as both are designed to limit certain anticompetitive forms of COLLUSION among would be rivals. One important distinction is that the EU law, unlike its US counterpart, allows the antitrust authorities to exempt some agreements. The conditions necessary for an exemption require that the agreement:

1 promotes TECHNICAL EFFICIENCY in the production or distribution of the good, and allows consumers a fair share of the benefits;
2 contains only restrictions which are indispensable to achieving the goal; and
3 does not allow the possibility of eliminating a substantial part of the competition.

Thus, for example, while market sharing is illegal *per se* in the US, there may be circumstances (e.g., among only a few, small firms) wherein this activity may be allowed in the EU.

Article 82 outlaws "abuse by one or more undertakings of a dominant position" which affects trade between EU member states. This law parallels (somewhat less closely) Section 2 of the Sherman Act, which proscribes any "attempt to monopolize" a market (*see* MONOPOLY). The law does not specify a minimum market share necessary for dominance, but Horspool and Korah (1992) and others cite cases where a share of 40 to 45 percent, combined with a lack of close rivals, has been identified by the courts as meeting the test. Moreover, the courts have indicated that the ability to act

independently from one's rivals (e.g., with re spect to pricing) provides sufficient evidence of dominance.

The law does identify some specific forms of conduct which constitute "abuse." The first cited is "unfair purchase or selling prices or other unfair trading conditions." This may in clude prices seen as excessively *high* to con sumers, or *low* prices which are seen as predatory to rivals. Second, limiting production or restricting product development is seen as abusive to consumer interests. A third case of abuse occurs when a dominant firm creates a competitive disadvantage for some trading part ners (e.g., suppliers) by discriminating (on price or other conditions) on the purchase or distribu tion of its products.

Overall the standards for dominance are stricter than in the US, where a minimum market share threshold closer to 65 percent has been established by the courts. In addition, the abusive anticompetitive conduct requirements are broader as well. Niels and ten Kate (2000), in a survey of several nations, argue that there is a convergence of standards concerning PREDA TORY PRICING among authorities in North America, who in general are suspect of its wide spread feasibility. As such, the legal requirement for finding predation in the US and Canada includes finding that high BARRIERS TO ENTRY and other MARKET STRUCTURE pre conditions are in place. In addition US courts typically require evidence of prices being set by the predator below its own variable costs. In contrast, Niels and ten Kate maintain that EU courts find aggressive price cutting in the face of new ENTRY may be abusive even if prices remain above costs, if the pricing creates a cred ible threat (*see* CREDIBLE STRATEGIES) in the eyes of potential rivals and thus reduces the level of competition in the future.

Articles 81 and 82 of the Treaty apply to competition that substantially crosses the na tional borders of EU members. Authority for writing and modifying this law lies primarily with the European Council of Ministers. While competition that primarily impacts a single nation has often been governed by that nation's laws, EU members have made a concerted effort to modify national laws which are inconsistent with those of the Union. Moreover, prior to

2004, the European Commission, which is the regulatory arm of the EU, has maintained the exclusive authority to grant exemptions to the law.

In Council Regulation (EC) No. 1/2003, the Council of Ministers adopted new rules on imple menting Articles 81 and 82, which went into effect in May 2004 (see European Economic Community Council, 2003, for details). Regula tion 1/2003 requires close cooperation between the member states and the Commission. In add ition to retaining the authority to end an in fringement of Articles 81 and 82, individual states may now also grant exemptions to Article 81. This offers national competition authorities and their courts greater powers to examine anti trust violations. It also removes an increasing burden of the European Commission – to exam ine and comment on national proceedings in volving possible exemptions – and frees its resources to focus on international violations and merger control. This additional empower ment for member states is tempered by the req uirement that they (1) are required to inform the Commission immediately after beginning an in vestigation, (2) are relieved of their authority if the Commission begins its own proceedings, and (3) are not allowed to take decisions which are counter to those already adopted by the Com mission.

Regulation 17, adopted in 1962, has provided the Commission with strong powers to enforce EU law by investigating potential violations. Managers should be aware that surprise inspec tions of a firm's financial and legal documents can be made at the discretion of the head of the Commission. These powers are extraordinary relative to US and many other nations' antitrust authorities, who would be required to demon strate to a judge that there was probable cause of a crime being committed prior to receiving a search warrant from the court. In addition, em ployees who are questioned do not immediately have the right to an attorney under current EU law. Regulation 1/2003, which replaces Regula tion 17, has strengthened the Commission's powers in that inspections will be allowed at the homes of company directors, management, or staff if a reasonable suspicion exists that rele vant records may be found there. In addition, management or staff may be interrogated by

investigators, and fines of up to 1 percent of the firm's annual TOTAL REVENUE may be imposed for intentionally supplying incorrect or misleading information.

The Commission also prosecutes cases. It may issue remedies such as cease and desist orders. It may impose fines which can reach up to 10 percent of a firm's worldwide annual total revenue for severe and lengthy violations of Articles 81 and 82. Moreover, Regulation 1/2003 allows for any structural or behavioral remedies necessary to effectively end the violation. Here, the imposition of structural relief is limited to cases where there are no equally effective behavioral remedies (*see* STRUCTURE CONDUCT PERFORMANCE PARADIGM). Finally, the Commission may impose daily penalties of up to 5 percent of the average daily revenues of a firm in order to compel the firm to end an infringement, comply with a binding decision, or submit to an inspection.

Much like the US Federal Trade Commission, which maintains an internal quasi judicial system whose decisions can be appealed to US federal courts, final EC decisions may be appealed to the European Court of First Instance. Individuals or firms seeking damages from antitrust violations can also bring private actions, although these suits must be brought in the individual's national court.

In a case which parallels one recently held in the US, the European Commission is investigating the Microsoft Corporation for possible antitrust violations. While the case has been settled in the US (*see* MICROSOFT ANTITRUST CASE), and several remedies have been imposed (*see* MICROSOFT ANTITRUST CASE: REMEDIES), the EU continues to examine the firm's TYING of its operating system to its browser and audio visual software. In addition the Commission is concerned with Microsoft's possible abuse of dominant position through its conduct regarding software for large institutional network servers, and by its ongoing unwillingness to release programming information which is vital to software rivals who wish to provide new programs that are compatible with Microsoft's operating system. If these activities are found by the Commission to constitute violations of Article 82, then additional remedies beyond those of the US Department of Justice may be imposed. Preliminary

indications are that the Commission may be willing to allow industry suppliers and consumers, in conjunction with Microsoft, to find amenable remedies, rather than to impose one itself. Kanter (2004) reports that the Commission is considering allowing PC manufacturers to choose what music media to install on Windows, rather than requiring that Microsoft sell an unbundled version of Windows (*see* BUNDLING). The Commission's decision and proposed remedies, likely to be handed down in 2004, will further clarify the extent to which EU and US antitrust policy differs.

UK ANTITRUST POLICY

The 1998 Competition Act substantially rewrote the law on competition within Great Britain, closely aligning UK law with that of the EU. As such it can be used to further illustrate some of the features of EU policy and also some differences which remain in member states' national policies. Chapters 1 and 2 of the new UK law are replicas of Articles 81 and 82 of EU policy. Utton (2000), in his analysis of this new law, cites several important features which constitute changes to earlier UK policy. First, there is a significant increase in the investigative and enforcement powers of the Director General of Fair Trading (DGFT), whose office evaluates competition and issues guidelines and exemptions to the law. The new law calls for the DGFT to prosecute cases, and also provides, for the first time, that fines of up to 10 percent of the firm's UK turnover may be levied for violations. Parallel to EU law, the DGFT may enter firms and examine their documents, but the UK law also allows for up to two years of jail time for obstruction of an investigation.

In addition, under UK guidelines market share is seen as important, but alone not determinative, of market dominance by a firm. A market share of 40 percent or greater, when coupled with other market structure factors, and the ability to behave independently of its rivals, are the principal indicators of dominance by a firm. A market share of less than 25 percent is generally seen as a "safe harbor" from prosecution. This analysis, once again, closely conforms to that followed under EU law. Interestingly, the DGFT guidelines appear to step ahead of the EU in their use of the more

theoretical approach to delineate market bound aries. The UK guidelines use the hypothetical monopolist or SSNIP approach (*see* MARKET DEFINITION), paralleling US policy (*see* MERGER GUIDELINES, 1992–7). While the European Commission announced the accept ance of this approach in 1997, Camesasca and Van den Bergh (2002) report that its use by the Commission and the EU courts has been lag ging, as more traditional qualitative and quanti tative measures continue to be the primary method used.

The EU and UK laws also converge in that they both focus on the need for consumers to benefit from any cooperative agreement among rivals which is to be exempted, and in their recognition of efficiencies as a possible defense for such agreements (*see* EFFICIENCY). In add ition, both laws maintain a consistent focus on excessive prices and price discrimination constituting evidence of abuse of power. Finally, the UK is currently examining the exclusions and exemptions to its law, in an effort to more closely conform to the recent EU regulatory changes.

Under the new UK law, the Mergers and Monopolies Commission, in place since 1973, is replaced with the Competition Commission. Earlier law required mergers to be evaluated on the basis of their impact on the "public interest," which required examining several costs and benefits of the merger. The new law, as does the old, allows for efficiency considerations in defense of a proposed merger. In an examination of earlier merger cases, Weir (1993) found that the Commission was largely skeptical of the effi ciency claim, and that few cases were decided favorably strictly on that basis. The EU law concerning mergers has been significantly re written for 2004. In addition, a newly adopted set of merger guidelines for business has been established. Each includes several factors, in cluding efficiency, which are to be considered in evaluating a merger. As UK and EU policy here is relatively new, the various positions that the authorities and their courts will ultimately take is yet to be fully determined.

See also *antitrust policy (US); capitalism; EU merger guidelines, 2004; EU merger policy, 2004*

Bibliography

Bergeron, J. (2001). Antitrust federalism in the European Union after the modernization initiative. *Antitrust Bul letin*, **46** (3), 513 35.

Camesasca, P. and Van den Bergh, R. (2002). Achilles uncovered: Revisiting the European Commission's 1997 market definition notice. *Antitrust Bulletin*, **47** (1), 143 86.

Church, J. and Ware, R. (2000). *Industrial Organization: A Strategic Approach*. Boston: McGraw-Hill Irwin.

European Communities (1997). Treaty of Amsterdam Amending the Treaty on European Union, the Treaties Establishing the European Communities and Related Acts. *Official Journal of the European Communities*, C 340, 10.

European Economic Community Council (2003). Council Regulation (EC) No. 1/2003 on the implementation of the rules on competition laid down in Articles 81 and 82 of the Treaty. *Official Journal of the European Commu nities*.

Horspool, M. and Korah, V. (1992). Competition. *Anti trust Bulletin*, **37**, 337 85.

Kanter, J. (2004). EU may demand Microsoft let PC makers pick media players. *Wall Street Journal*, B3.

Martin, S. (1994). *Industrial Economics*, 2nd edn. New York: Macmillan.

Niels, G. and ten Kate, A. (2000). Predatory pricing standards: Is there a growing international consensus? *Antitrust Bulletin*, **45**, 787 809.

Palim, M. (1998). The worldwide growth of competition law: An empirical analysis. *Antitrust Bulletin*, **43**, 105 45.

Shishkin, P. (2002). Tough tactics: European regulators spark controversy with "dawn raids." *Wall Street Jour nal*, **239** (42), 1 6.

Utton, M. (2000). Going European: Britain's new com petition law. *Antitrust Bulletin*, **45**, 531 51.

Weir, C. (1993). The implementation of merger policy in the UK 1984 1990. *Antitrust Bulletin*, **38**, 943 67.

EU merger guidelines, 2004

Gilbert Becker

In May of 2004, the European Union (EU) adopted new law, Regulation No. 139, for the control of mergers in the European Community (*see* EU MERGER POLICY, 2004). The Euro pean Commission (EC) concurrently issued a set of guidelines in order to provide transparency and clarity to the business community as to the nature of the regulatory changes being made and

the process by which horizontal mergers will be examined in the future. The guidelines outline the potential economic problems and benefits arising from horizontal mergers in general, and detail the means by which the competitive con sequences of a merger are to be measured and weighed (see EC Notice, 2004, for details). Regulation 139 concerns "concentrations bet ween undertakings" and also applies to vertical and conglomerate mergers as well as joint ven tures among firms. Guidelines for these areas are forthcoming.

For several reasons the merger guidelines are particularly important to business executives in the EU. First, the EU decision making process is administrative rather than judicial. The EC evaluates proposed mergers and may by its own order block a merger without initial judicial app roval from the European courts, if it considers the proposed merger to have significant antic ompetitive effects. Under US law (see CLAY TON ACT; MERGER GUIDELINES, 1992–7) the antitrust authorities may evaluate a proposed merger, but must seek judicial action to have a merger stopped. Although in practical terms the US government's threat to challenge a merger may effectively thwart many mergers, the strength of the EC's authority further enhances the need for management's awareness of the Commission's method of applying the EU merger control regulation. Second, the guide lines indirectly offer insight as to possible defen ses which may be put forward in order to promote questionable mergers. Finally, for mergers involving significant commercial activ ity on both continents, US and EU authorities have agreed to a simultaneous review process, but each has retained its authority to evaluate a merger according to its own rules and economic philosophy.

The EU guidelines recognize that Regulation 139 maintains a focus on whether or not a merger would be likely to significantly impede effective COMPETITION (SIEC), especially through the creation or strengthening of a dominant pos ition. The guidelines further recognize the goal of maintaining effective competition in markets in order to bring benefits to consumers, includ ing low prices, high quality, a variety of choices of goods, and innovation. As such the review process involves two important steps:

1   defining the relevant market of the merging firms (see MARKET DEFINITION) and the structure of the market (see MARKET STRUCTURE); and
2   assessing the merger's impact on competi tion in the market.

An earlier notice issued by the EC (see EC Notice, 1997) explains that the Commission's method for defining the relevant economic market includes delineating a relevant product market and relevant GEOGRAPHIC MARKET and involves "identifying the effective alterna tive sources of supply for customers." The EC definition of a market rests primarily on demand substitutability, as a firm's ability to influence price is seen as being constrained most directly by consumers' ability to readily switch to alter native SUBSTITUTES. Supply side substitut ability, especially when rivals would be required to significantly adjust their assets, and the potential entry of new firms, are considered by EU regulators at a later stage of the merger review.

The EC considers a variety of evidence in its efforts to define a market. The Commission considers similarity in products' characteristics and in the intended use of the goods as evidence of their substitutability, but relies more heavily on its own surveys of customers and their views of the market's boundaries. Empirical examin ations showing a similarity of prices and a cor relation of price movements of rival products, as well as measures of significant cross elasticity of demand (see CROSS ELASTICITIES), also pro vide evidence that goods are in the same market. With respect to finding the geographic market boundary, the Commission examines evidence of the importance of national or local preferences and the records of past shipments of products across regions, among other factors.

The Commission's use of many sources of evidence is consistent with the practices of both the US and EU judiciary. In addition, both US and EU antitrust authorities recognize the im portance of a relatively new approach in which a hypothetical monopolist offers a small but significant nontransitory increase in prices (SSNIP) and the market's boundaries are deter mined by the reactions of buyers and sellers (see the entry on market definition for a full explan

ation). In defining a market, US and EU practices diverge in that US regulators focus more attention on supply side substitutability of goods, and rival firms' reactions to a SSNIP. The methods contrast further in that the EU approach, especially in phase 1 of an investigation, seeks to discover whether a reasonable, yet perhaps overly narrow, definition of the relevant market yields no competitive concerns, thereby ending the inquiry within the new regulation's limited time frame. The more open ended time frame of the US process encourages a more thorough yet time consuming initial analysis, in an effort to "get the definition right" in the first place.

The remaining elements of the new EU guidelines largely converge with their US counterpart. Paralleling the US approach, market concentration is examined using the HERFINDAHL HIRSCHMAN INDEX (HHI), and three levels of concentration are defined. Markets with a post merger HHI measure of less than 1,000 are unlikely to be challenged, as are those where the post merger HHI falls between 1,000 and 2,000, and the change in HHI is less than 250. Markets where the post merger HHI exceeds 2,000 are also unlikely to raise EC concerns if the rise in HHI is less than 150. These standards are slightly more lenient than those in the US, but the EU guidelines are careful to note the HHI levels are only initial indicators, and are not presumptive evidence of the presence or absence of competitive concerns. Also, special circumstances such as the merger involving a potential rival, an important innovator, or an industry maverick may effectively lower these thresholds. Unlike the US guidelines, the EC also examines the leading firm's market share. Here a market share below 25 percent ordinarily constitutes a safe harbor for the merging firms, while anything above 40 percent may give rise to competitive concerns and further analysis.

The guidelines recognize that a merger may significantly impede effective competition in two ways. The first is through so called unilateral effects (UE) in which a merger may remove competitive constraints on a leading firm (or group of firms in an OLIGOPOLY). Here a dominant position may be created or strengthened without the coordination of the remaining rivals. A merger among firms which are already large may result in the newly created entity having a sufficiently large market share that there is an increased likelihood of enhanced MARKET POWER. Moreover, if prior to the merger a significant percentage of the customers of the two merging firms ranked their products as the top two choices, the ability of the newly merged firm to raise its prices may be strengthened. The guidelines identify several other factors including high consumer SWITCHING COSTS (see LOCK IN; TRANSACTIONS COSTS) and the removal of an aggressive rival as other circumstances in which consumers may have greater exposure to UE harm.

Coordinated effects (CE) arise when a merger results in the increased ability of the remaining firms to coordinate their pricing, output, or other business strategies (see COLLUSION). This interfirm cooperation is more likely to occur in oligopolistic markets, and the guidelines follow conventional microeconomic theory (see CARTELS; GAME THEORY) in presenting numerous factors which may enhance the ability to coordinate, monitor, and enforce an agreement, as well as to diminish the impact of rivals who remain outside the agreement. Where such factors exist, a merger yielding coordinated effects may significantly impede effective competition and be assessed as being incompatible with the common market.

In further evaluating the competitive effects of a proposed merger the Commission will examine (1) countervailing buyer power, (2) entry into the market, and (3) efficiencies. Organized buyers can sometimes provide a credible threat of switching to an alternative good in response to a newly merged firm's price increase. Here the guidelines offer a possible defense for an objectionable merger. If the buyer's countervailing power can be shown to broadly protect consumers, and remain effective after the merger, then the risk of the merger significantly impeding effective competition is reduced. The new EU guidelines, following those in the US, also examine the question of new ENTRY into the market (see BARRIERS TO ENTRY). Here, the demonstration that new entry is *likely* to occur in a *timely* fashion and with *sufficient* magnitude to thwart any anticompetitive unilateral or coordinated effects may also provide a defense on behalf of a proposed merger.

Finally, the guidelines recognize that mergers may offer pro competitive effects in light of efficiencies which may be achieved by the com bination of firms (*see* EFFICIENCY; TECH NICAL EFFICIENCY). The guidelines offer that the presence of efficiencies may provide grounds for declaring a merger compatible with Regulation 139. Once again the EU and US guidelines converge, as both require that the efficiencies are (1) merger specific (may only be achieved by the merger); (2) verifiable; and (3) likely to cause the merged firm to provide benefits to consumers, such as lower prices or new or improved products.

The EU guidelines indicate that the above defenses depend on the particular circumstances of each merger. Brandenberger and Gudofsky (2004) report that the Commission historically has been receptive to an entry defense, and some what amenable to the buyer power defense. They indicate that, with some exceptions, the EC has been much more hostile to an efficiencies de fense. In particular, production economies (*see* ECONOMIES OF SCALE), by providing a means for achieving or strengthening dominance, have at times been treated as if they were an antitrust *offense*. The new guidelines, and the EC merger regulation itself, provide language indicating that efficiencies will be treated differently in the future, and may provide support for a proposed merger. As the new law is enforced and guide lines are subjected to a number of cases, their interpretation and application will be revealed.

See also *merger guidelines, 1992–7*

Bibliography

Brandenberger, R. (2004). EU merger control: Develop ments in 2003. In *US and EU Antitrust Aspects of Mergers and Acquisitions*. Los Angeles: UCLA School of Law.

Brandenberger, R. and Gudofsky, J. (2004). Potential defenses in EU merger control. In *US and EU Antitrust Aspects of Mergers and Acquisitions*. Los Angeles: UCLA School of Law.

Commission of the European Communities (1997). Com mission Notice on the definition of the relevant market for the purposes of Community competition law. *Offi cial Journal of the European Union*, C372.

Commission of the European Communities (2004). Com mission Notice: Guidelines on the assessment of horizontal mergers under the Council Regulation on the control of concentrations between undertakings. *Official Journal of the European Union*, C31/5.

Gotts, I. K. and Goldman, C. S. (2002). *The Role of Efficiencies in M&A Global Antitrust Review*. New York: Fordham Corporate Law Institute.

Levy, N. (2003). EU merger control: From birth to ado lescence. *World Competition*, **26** (2).

Van Kerckhove, M. and Ersboll, N. (2004). Co-ordinated interactions: Background paper on EC Case Law and the Commission's Notice on the assessment of horizon tal mergers. In *US and EU Antitrust Aspects of Mergers and Acquisitions*. Los Angeles: UCLA School of Law.

Volcker, S. (2004). Mind the gap: Unilateral effects an alysis arrives in EC merger control. In *US and EU Antitrust Aspects of Mergers and Acquisitions*. Los Angeles: UCLA School of Law.

EU merger policy, 2004

Gilbert Becker

In 2004 the European Union (EU) revised its regulation concerning mergers between firms which produce goods or services in the EU. The European Council, in establishing Regula tion No. 139/2004 (the EC Merger Regulation), recognized that the 2004 enlargement of the EU and the continuing decline in trade barriers be tween states would result in additional reorgan izations of firms. As such, the regulation cites as one of its objectives the facilitation of this eco nomic integration, when it promotes COMPETI TION and enhances EU standards of living (see EU Council, 2004, for a complete statement of the regulation).

Following a series of stinging rebukes from the Court of First Instance (see Brandenburger, 2004, for details), the European Commission (EC), the regulatory branch of the EU, reorgan ized its merger decision making structure, en hanced its economic analysis, and produced a complementary set of guidelines for evaluating horizontal mergers. The latter were designed to help the EU business community to understand the new regulation being adopted and the cir cumstances under which mergers would be pro scribed (*see* EU MERGER GUIDELINES, 2004).

In an effort to strengthen the industry know ledge of those who will review mergers, begin ning in 2004 the EC's Merger Task Force has

been integrated into the four existing antitrust directorates in the Competition Directorate General, each of which specializes in one sector of the economy including (1) information and communications, (2) industry and energy, (3) services, and (4) capital and consumer goods. In addition, a Chief Competition Econo mist position has been established and a staff of ten economists will provide an independent evaluation of proposed mergers, along with the examination of other potential violations of anti trust policy.

New procedures in the Merger Regulation also promote a more streamlined and flexible review process. In an effort to expedite the pro cess, merging firms, who remain required to notify authorities prior to the consummation of a merger, may now pre notify the Commission in cases where an agreement is imminent. The review itself comes in two stages. The phase 1 preliminary investigation is limited to 25 days. During this phase the firms are required to provide detailed information and to establish an initial position on their view of the relevant market (*see* GEOGRAPHIC MARKET). If the merger is not approved, a more in depth phase 2 investigation may be undertaken but is limited to 90 days (slightly more if remedies are offered), unless the firms themselves "stop the clock" in order to gather more information in their def ense. Finally, the new regulation creates a more transparent process, in part by requiring the Commission to offer, in writing, an explanation for the rejection of a proposed merger.

The principal change embodied in Regulation 139 stems from the language within Articles 81 and 82 of the EU Treaty (*see* EU COMPETITION POLICY, 2004) and provides a somewhat new foundation for the analysis of mergers. Under the previous (1997) merger regulation, the cen tral question to be resolved was whether the merger was likely to *create or strengthen a domin ant position*. This standard has been modified and the language of the regulation changed to now invalidate mergers which would "signifi cantly impede effective competition" (SIEC). Although the new language has not yet been interpreted in the EU courts, and its effects have yet to be fully determined, one significant loophole is likely to be closed by the regulation. The flaw in the earlier regulation exists in the

case of a proposed merger between the second and third largest firms in a market. If this com bination would not achieve dominance through the merger, it may ordinarily have been allowed under the previous EU regulation, and indeed such a merger may enhance competition in the relevant market(s) by providing a more formid able rival to the leading firm. The new SIEC test recognizes the possibility that this merger may reduce competition by facilitating COLLUSION among the rivals in an oligopolistic market (*see* OLIGOPOLY). This may result from the reduc tion in the number of competitors, facilitating both the initial organization and the future com munication among those in the market, or by improving the ease of enforcing the cooperative behavior required for their collective success (*see* CARTELS; STRUCTURE CONDUCT PER FORMANCE PARADIGM). These circumstances may be further enhanced in cases where the merger involves the acquisition of a particularly aggressive competitor, whose removal may sig nificantly change the competitive environment in the market.

The new regulation continues to cite the strengthening and possible abuse of a dominant position as the primary basis for rejecting a pro posed merger under the SIEC standard. As such, the scope of the regulation is now broadened in that it includes the examination of the *coordinated effects* of the merger, i.e., enhanced likelihood of cartel like behavior, as well as the possible *uni lateral effects* of the merger emanating from the enhanced dominance of a single firm.

EU law regarding mergers is now more closely aligned with that of the US. The new SIEC language is a close approximation to that of the US CLAYTON ACT, which condemns mergers that "substantially lessen competition." In add ition, the EU regulation's continued concern with "abuse of dominance" maintains a parallel to the Clayton Act's proscription of mergers which "tend toward monopoly" (*see* MONOP OLY). Moreover, the future application of the two laws, through the use of their respective guidelines, indicates further convergence. In fact the US guidelines, which make specific ref erences to both the "unilateral" and "coordin ated" effects of mergers, have provided a framework for those written in the EU (see MERGER GUIDELINES, 1992–7).

This convergence is particularly important to multinational firms and those with intentions to expand their market activity through inter national mergers. While the application of earlier EU and US law had often drawn similar conclu sions concerning the competitive effects of a merger, an important exception occurred in 2001 when the proposed merger of General Electric Co. and Honeywell Inc. was first ap proved by US antitrust authorities, but subse quently stopped by the EC which found that the proposed merger would have had unacceptable anticompetitive effects in the EU.

While this convergence continues, there remain differences between the approaches of the two unions. First, while both laws offer the opportunity to investigate and condemn vertical and conglomerate mergers, the EU currently appears to be somewhat more active and strict in these areas. Second, the EU regulation pro vides direct lines for input by its member states into the EU's investigation and evaluation of mergers. Third, and perhaps most importantly, the EU process grants the Commission the au thority to prohibit mergers, while the US system allows its authorities to investigate proposed mergers, and threaten a judicial challenge, but requires a court injunction to legally stop a merger. Finally, Gifford and Kudrle (2003) argue that while convergence of EU and US antitrust policy in general continues, it is con strained by legal, political, and ideological tradition.

See also *antitrust policy (US); horizontal merger guidelines; merger guidelines, 1992–7*

Bibliography

Brandenberger, R. (2004). EU merger control: Develop-ments in 2003. In *US and EU Antitrust Aspects of Mergers and Acquisitions*. Los Angeles: UCLA School of Law.

Brandenberger, R. and Gudofsky, J. (2004). Potential defenses in EU merger control. In *US and EU Antitrust Aspects of Mergers and Acquisitions*. Los Angeles: UCLA School of Law.

Church, J. and Ware, R. (2000). *Industrial Organization: A Strategic Approach*. Boston: McGraw-Hill Irwin.

European Union Council (2004). Council Regulation No. 139/2004 on the control of concentrations between undertakings (the EC Merger Regulation). *Official Journal of the European Union*.

Gifford, D. and Kudrle, R. (2003). European Union competition law and policy: How much latitude for convergence with the United States? *Antitrust Bulletin*, 48 (3), 727 80.

Levy, N. (2003). EU merger control: From birth to ado-lescence. *World Competition*, 26 (2).

US Department of Justice and Federal Trade Commis-sion (April 2, 1992; revised April 8, 1997). *Horizontal Merger Guidelines*. Washington, DC: US Government Printing Office.

Volcker, S. (2004). Mind the gap: Unilateral effects an-alysis arrives in EC merger control. In *US and EU Antitrust Aspects of Mergers and Acquisitions*. Los Angeles: UCLA School of Law.

excess capacity

Nikolaos Vettas

A firm operates with excess capacity when its output is below the level allowed by its invested capacity, that is, below the level at which all its production inputs are fully employed. Despite the costs involved in this practice (primarily reflecting the opportunity cost of the inputs not utilized; *see* OPPORTUNITY COSTS), a firm may maintain excess capacity for several reasons. These reasons include demand fluctuations over time (due to either seasonality or the business cycle), and ECONOMIES OF SCALE when demand grows over time. Excess capacity may also be used as a strategic BARRIER TO ENTRY (see Salop, 1979). The argument is that excess capacity makes it less costly for an incum bent firm to increase sales, and, thus, increases the credibility of "aggressive" behavior (such as price cutting) after ENTRY occurs. The potential entrants recognize this prospect, and thus entry is deterred (*see* CREDIBLE STRAT EGIES).

Excess capacity has also been studied in the context of MONOPOLISTIC COMPETITION (introduced by Chamberlin, 1933). The basic question explored there is whether free entry yields too many firms from a social viewpoint, or if firms operate at too low a rate of output to exhaust economies of scale. Chamberlin argued that firms in monopolistically competitive markets would not operate at the minimum of their average cost curve and therefore excess capacity existed in these markets. This issue

has been further explored in more recent research (see Dixit and Stiglitz, 1977).

Bibliography

Chamberlin, E. H. (1933). *The Theory of Monopolistic Competition*. Cambridge, MA: Harvard University Press.

Dixit, A. and Stiglitz, J. (1977). Monopolistic competition and optimum product diversity. *American Economic Review*, **67**, 297 308.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Salop, S. (1979). Strategic entry deterrence. *American Economic Review*, **69**, 335 8.

exit

Robert E. McAuliffe

Firms exit an industry when they cease production, either voluntarily or through forced BANKRUPTCY proceedings. Exit may occur because an industry is declining or because of intensified competition between existing firms in the industry that forces some firms out of the market. In their study using US Census of Manufacturers data over the period from 1963 to 1982, Dunne, Roberts, and Samuelson (1988) found that industries with high rates of ENTRY also had high rates of exit. This suggests that industry specific factors may influence both types of behavior. The large numbers of firms entering and exiting manufacturing indicate significant activity and turnover in this sector. Those firms exiting an industry tended to be smaller on average than existing firms and, despite their large number, also tended to account for a small percentage of market output.

Ghemawat and Nalebuff (1985) developed a theoretical model of exit as a strategic game between firms in a market. Given that declining market demand will force reductions in an industry's productive capacity, they questioned whether small or larger firms would be the first to exit. When the firms had identical costs, the larger firms were predicted to leave the industry first because the smaller firm could operate profitably at lower output levels for a longer period of time relative to a large firm. Therefore, if both firms know that industry sales will decline, the small firm can wait for the large firm to exit first because the smaller firm can hold out longer, knowing that it will be more profitable in a DECLINING INDUSTRY than the larger firm. Lieberman (1990) called this prediction the "stakeout hypothesis." If the larger firm enjoyed cost advantages over the smaller firm, perhaps through ECONOMIES OF SCALE, this conclusion could be reversed. The interesting prediction from this model was that market concentration would decrease as the market declined.

This theoretical conclusion was later qualified by Whinston (1988) and Londregan (1990). Whinston argued that if a large firm were a multiplant producer, it could reduce its capacity by closing plants without necessarily exiting the industry. From his results, Whinston shows that no general predictions about the order of firm exit can be made when there are multiplant firms since the pattern of exit depends upon the specific characteristics of the industry. He also showed that the exits may not be best from society's perspective: the more efficient firms may exit first and/or the more efficient plants may be shut down first, leaving the market to the less efficient producers. This less efficient outcome could arise because the more efficient firms or plants might be those in the industry that are larger. Since they need more sales to be profitable (relative to smaller firms or plants), they cannot hold out against smaller firms in a declining market because the smaller firms will ultimately be more profitable as industry output falls. Thus the smaller firms can afford to wait and, knowing this, the larger firms will choose to exit first. Londregan (1990) showed that when firms compete over the INDUSTRY LIFE CYCLE and can reenter a market if they choose, high costs of entry act as a barrier to exit. Once a firm has incurred the startup costs necessary to enter the industry, it is less willing to exit the industry (and possibly have to incur the startup costs again). These entry costs can create advantages for less efficient producers who can preempt the market and keep more efficient firms from entering.

Lieberman (1990) tested the exit models to determine whether large firms were at a disadvantage when markets for particular products

declined in the chemical industry. In his sample of declining products, output declined 42 per cent on average from the industry production (or capacity) peak, while the number of producers fell 32 percent. The pattern of exit from the industry was consistent with the "shakeout" hypothesis in that smaller firms were much more likely to fail and exit the industry. However, the "stakeout" hypothesis of Ghemawat and Nalebuff (as amended by Whinston) was also supported because when there were steep declines in industry capacity, larger firms tended to reduce their capacity by a greater proportion relative to small firms and, in this respect, larger firms were at a disadvantage when industry demand fell. Lieberman suggests that exit patterns are affected by two offsetting forces: economies of scale, which create advantages for larger firms and increase their likelihood of survival, versus the strategic liability of being a large firm in competition with smaller firms when the market declines.

See also *declining industry*

Bibliography

Dunne, T., Roberts, M. J., and Samuelson, L. (1988). Patterns of entry and exit in US manufacturing industries. *The Rand Journal of Economics*, **19**, 495 515.

Ghemawat, P. and Nalebuff, B. (1985). Exit. *The Rand Journal of Economics*, **16**, 184 94.

Lieberman, M. B. (1990). Exit from declining industries: "Shakeout" or "stakeout"? *The Rand Journal of Economics*, **21**, 538 54.

Londregan, J. (1990). Entry and exit over the industry life cycle. *The Rand Journal of Economics*, **21**, 446 58.

Whinston, M. D. (1988). Exit with multiplant firms. *The Rand Journal of Economics*, **19**, 568 88.

expected present value

John Edmunds and Roberto Bonifaz

Future cash flows are sometimes known with certainty, for example, if they are coming from a portfolio of government bonds (*see* CASH FLOW). More often, however, there is some uncertainty around exactly what the future cash flows will be. The future cash flows from an investment during the next three years might be $100 each year, but could be lower or higher, depending on how future events develop. The uncertainty can be described in terms of probability distributions, which could be discrete or continuous. For example, there might be a probability of 0.8 that the cash flow for year 1 will fall between $90 and $110. The probabilities might be symmetric around the most likely outcome, or they might be skewed. Whether the probabilities are symmetric or skewed, the expected value of each year's cash flow can be computed. This is done by assigning a probability to each possible amount of cash flow that might occur, then proceeding to multiply each outcome by its probability of occurring and adding the products together.

In mathematical terms for the discrete case:

$$\text{expected cash flow} = \sum_i x_i p(x_i)$$

where x_i represents each possible cash flow outcome, $p(x_i)$ is the density function associated with the outcomes, and the summation is over all i outcomes.

In the continuous case the expected cash flow is expressed as:

$$\text{expected cash flow} = \int x f(x) dx$$

where x is the random variable describing future cash flows and $f(x)$ is the corresponding density function.

With the expected future cash flows computed, it is possible to compute expected present value. Each expected future cash flow is discounted, giving its present value. The present values of all the expected future cash flows are summed, yielding the total expected present value. If the initial investment is subtracted from the total expected present value, the resulting number is called expected net present value.

In mathematical terms, for the discrete case:

$$\text{expected present value}$$
$$= \sum_j \left(\sum_i x_i p(x_i) \right) (1+r)^{-j}$$

where j are the time periods and the summation is over all outcomes, i.

In the continuous case:

$$\text{expected present value} = \iint e^{-rt} x f(x)\,dx\,dt$$

where t represents the continuous passage of time.

The naive decision rule is that a firm should approve every investment project that offers a positive expected net present value. This rule ignores RISK AVERSION (see Friedman and Savage, 1948), and it also excludes the necessity of scrutinizing the distribution of the expected cash flows. An investment with a positive expected net present value has a measurable probability of giving a negative return. It is an essential part of the analysis to compute the probability that the return on investment will be negative. Business managers also calculate the probability that the return on investment will be lower than some minimum acceptable amount. They express this as a rate of return and call it the hurdle rate. If each of the annual cash flows comes out in the lower tail of its probability distribution, the final outcome of the investment will be worse than its expected present value.

Bibliography

Brigham, E. F. and Gapenski, L. C. (1997). *Intermediate Financial Management*, 5th edn. New York: International Thomson Publishing.

Friedman, M. and Savage, L. J. (1948). The utility analysis of choices involving risk. *Journal of Political Economy*, **56**, 279–304.

expected value

Robert E. McAuliffe

The mean, or expected value, is a measure of central tendency for a random variable and it is one of the characteristics of probability distributions (the VARIANCE is another common characteristic). If the random variable, X, is continuous and has specific values denoted by x, and the probability distribution of the values of X is given by $f(x)$, then the expected value of X, $\text{E}(X)$, is calculated as

$$\text{E}(X) = \int_X x f(x)\,dx \qquad (1)$$

where $\int$ is the integral over the complete range of *possible* values for X (see Greene, 2002; Kmenta, 1986). When the random variable X is discrete, the expected value is calculated as

$$\text{E}(X) = \sum_x x f(x) \qquad (2)$$

where $\sum_x$ indicates the summation is over all possible values of x.

In business analysis, managers work with observed sample data and not necessarily with formal probability distributions. The sample mean, denoted $\bar{X}$, can then be calculated for the observed values of the variable of interest. For example, a firm may want to calculate its average weekly sales in the last quarter from weekly data. In this case there are 12 weekly observations and the sample mean would be calculated by simply summing the weekly sales figures and dividing by 12. If a manager randomly selected any single week from the 12, any single week's sales would be equally likely, so the probability of any specific week's sales being selected would be one twelfth and this is the value $f(x)$ would take in the summation in (2) above. The summation in equation (2) would then be:

$$\bar{S} = \sum_{i=1}^{12} S_i \times \frac{1}{12} = \frac{1}{12} \sum_{i=1}^{12} S_i \qquad (3)$$

where S_i is the level of sales in each week. It should be noted that the sample mean of sales calculated above may not equal any of the sales figures which occurred, but rather it represents the likely value for weekly sales if the company were to have several quarters like the last. Suppose, for example, that sales were 1 million units for the first six weeks of the quarter and 2 million units for the second six weeks. The mean value of sales would then be 1.5 million units, a value which did not occur during the sample period.

For forecasting, a manager will need to assign probabilities based on her best estimate of the likely outcomes since past data will not be available or may not apply. In these cases it is crucial to perform sensitivity analysis to determine how sensitive the conclusions or forecasts are to changes in the underlying assumptions (or probabilities).

Bibliography

Greene, W. H. (2002). *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.

Gujarati, D. N. (2002). *Basic Econometrics*, 4th edn. New York: McGraw-Hill Irwin.

Kmenta, J. (1986). *Elements of Econometrics*, 2nd edn. New York: Macmillan.

Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.

experience goods

Robert E. McAuliffe

Experience goods are those products which must be consumed or used by consumers before their quality can be determined, such as automobiles, food, and software. Nelson (1970, 1974) distin guished between these products and SEARCH GOODS, and he suggested that the information provided by ADVERTISING would depend upon whether the products were search goods or exp erience goods. According to Nelson, advertising for experience goods would rely on imagery more than factual information and would adver tise proportionately more often in broadcast media than in print media. Furthermore, since consumers cannot determine the quality of exp erience goods without purchasing them and ad vertising cannot convey much direct information about the product, Nelson argued that it is the *volume* of advertising which conveys information about product QUALITY to consumers. Firms that sell higher quality goods or which are more efficient in producing the good have greater incentives to advertise because the profits from additional sales will be greater. Higher quality firms will benefit from repeat purchases in the future if consumers try their products, and more efficient firms will earn a greater profit per unit sold than other firms which are less efficient or produce lower quality products. Therefore Nelson claimed that firms which produce higher quality experience goods have greater in centives to advertise and, as they do so, con sumers will use information about the volume of advertising to infer product quality.

Nelson's model offers several hypotheses: exp erience goods will be advertised more than search goods; experience goods are likely to be advertised in media which provide "soft" infor mation rather than the "hard," factual informa tion provided for search goods; products which are purchased frequently will be advertised more than products which are purchased less fre quently; and durable goods will be advertised less than nondurable goods (whether they are search or experience goods) because they are purchased less frequently. Nelson found search products tended to be advertised more fre quently in newspapers and magazines than exp erience goods. In addition, advertising in local media should provide more relevant information than advertising in national media so experience goods should have a greater proportion of their advertising expenditures in national media, a prediction which was also supported by his data. Nelson (1974) also found a higher advertis ing–sales ratio for nondurable goods than for durable goods. Ferguson (1974) observed that the distinction between search and experience goods can be somewhat arbitrary, yet the distinc tion is crucial to his theory and tests. Ehrlich and Fisher (1982) extended the model, and a survey and tests of the advertising–profitability issue can be found in McAuliffe (1987). Managers can use Nelson's categories to guide their adver tising decisions with respect to the volume, con tent, and appropriate media for their products.

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Indus trial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Ehrlich, I. and Fisher, L. (1982). The derived demand for advertising: A theoretical and empirical investigation. *American Economic Review*, **72**, 366 88.

Ferguson, J. M. (1974). *Advertising and Competition: Theory, Measurement, Fact*. Cambridge, MA: Ballinger.

McAuliffe, R. E. (1987). *Advertising, Competition, and Public Policy: Theories and New Evidence*. Lexington, MA: D. C. Heath.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, **78**, 311 29.

Nelson, P. (1974). Advertising as information. *Journal of Political Economy*, **82**, 729 54.

externalities

Mark Rider

Externalities are the costs or benefits from activ ities engaged in by one set of economic agents that affect other economic agents who are not directly engaged in the activity. For example, the noise from aircraft overhead diminishes the util ity (*see* TOTAL UTILITY) or happiness of indi viduals living in areas surrounding an airport. The diminished utility resulting from noisy air craft is an external cost of production. Further more, external costs or benefits may occur in either the production or consumption of a good, or both. An example of an external benefit in consumption is the pleasure enjoyed by the neighbors of an individual who plants a flower bed. Generally, goods or activities with external benefits are called public goods (*see* PUBLIC GOODS PROBLEM), and the term externality typically refers to an external cost. Other common examples of externalities are industrial pollution, automobile emissions, and traffic con gestion.

External costs or benefits prevent markets from achieving an efficient allocation of re sources (*see* EFFICIENCY). The market will al locate too many resources to the production of goods with external costs and too few to the production of goods with external benefits. For example, ticket prices will not equal the true MARGINAL COST if airlines do not take into account the harm imposed on others by their noisy aircraft. Consequently, more resources will be devoted to air travel than in an efficient allocation.

Externalities may require a public sector re sponse in order to remedy the misallocation of resources. There are three tools for controlling externalities: regulation, property rights, and taxes. Frequently, governments issue extensive regulations requiring industry to reduce the ext ernal costs associated with certain activities. An important instance of the federal government using regulation to control an externality is the automobile emission standards contained in the Clean Air Act.

Command and control type regulation lacks the necessary flexibility for individuals and com panies to find more efficient methods of redu cing external costs. However, government can get companies to internalize external costs in their decision making by properly establishing and enforcing property rights (*see* COASE THE OREM). For example, the government could issue vouchers allowing the holder to pollute a given amount. Companies that can reduce emis sions relatively inexpensively could sell their vouchers to companies for whom it is more costly to reduce emissions. In this manner, gov ernment could achieve the desired reduction in pollution with lower OPPORTUNITY COSTS to society than under a command and control regime. Furthermore, tradeable vouchers create incentives for firms to strive continuously to lower the cost of reducing pollution. Alterna tively, the government could levy taxes on com panies equal to the external costs associated with their activities. Again, firms would internalize the externalities and are thereby motivated to reduce the externality in the most cost effective manner.

Bibliography

Baumol, W. J. (1972). On taxation and the control of externalities. *American Economic Review*, **62**, 307 22.

DeSerpa, A. (1978). Congestion, pollution, and impure public goods. *Public Finance*, **33**, 68 83.

Oates, W. E. (1983). The regulation of externalities: Effi cient behavior by sources and victims. *Public Finance*, **38**, 362 75.

F

factor prices

Laura Power

Factor prices are the costs (i.e., prices) associated with inputs (i.e., factors) used in production. The primary factors of production include land, labor, and capital. In competitive markets, the price for each of these factors depends on the EQUILIBRIUM between supply and demand. Total demand for a particular factor is deter mined by the combination of all individual firms' demands for that factor, which in turn are determined by each firm's PROFIT MAXI MIZATION decisions. The determination of total supply is factor specific; for example, the supply of land is relatively fixed, whereas the supply of labor depends on each individual's willingness to work (UTILITY MAXIMIZA TION).

The equilibrium pricing of factors serves to allocate resources efficiently (*see* EFFICIENCY). In markets where land is scarce and labor plenti ful, the price of land is high, and that of labor is low; whereas in markets where land is plentiful and labor scarce, the reverse is true. At the international level, however, if free trade exists, and if several restrictive assumptions are ful filled, differences in factor prices across coun tries hypothetically equalize. This theory is based on the Heckscher Ohlin approach to inter national trade, and is called factor price equal ization.

Bibliography

Albrecht, W. P. (1983). *Economics*, 3rd edn. Englewood Cliffs, NJ: Prentice-Hall.
Samuelson, P. A. and Nordhaus, W. D. (2001). *Economics*, 17th edn. New York: McGraw-Hill Irwin.

Varian, H. R. (1992). *Microeconomic Analysis*, 3rd edn. New York: Norton.

factor productivity

Laura Power

Factor productivity is defined as the output per unit of factor input. The primary factors of production include labor, land, and capital. At the micro level, factor productivities depend on each firm's optimal input combination, which is determined by its PROFIT MAXIMIZATION decision. The level of output associated with the optimal choice of inputs is determined by the production function (*see* PRODUCTION FUNCTIONS), and the ratios of this output to each input are the factor productivities for the firm.

In markets where the price of labor is high relative to that of capital, firms will tend to employ more capital (i.e., will choose high cap ital–labor ratios), relative to markets in which the price of labor is low relative to that of capital. Firms with such high capital to labor ratios are called capital intensive firms, while firms with high labor to capital ratios are called labor inten sive firms. Capital intensive firms have high output per unit of labor input (labor productiv ity) relative to labor intensive firms.

See also *efficiency; factor prices; marginal prod uct*

Bibliography

Albrecht, W. P. (1983). *Economics*, 3rd edn. Englewood Cliffs, NJ: Prentice-Hall.

Samuelson, P. A. and Nordhaus, W. D. (2001). *Economics*, 17th edn. New York: McGraw-Hill Irwin.

Varian, H. R. (1992). *Microeconomic Analysis*, 3rd edn. New York: Norton.

Federal Trade Commission Act

Gilbert Becker

The Federal Trade Commission Act is the third principal US antitrust statute (*see* CLAYTON ACT; SHERMAN ACT). This statute created the Federal Trade Commission (FTC) and empowered it to enforce the Clayton and FTC Acts. The primary section of the FTC Act, Section 5, proscribes "unfair methods of competition" and "unfair or deceptive acts or practices in or affecting commerce." Unlike the Sherman and Clayton Acts, over which the US Department of Justice also holds jurisdiction, and through which private individuals or firms may also file suit, the FTC Act gives the Commission exclusive authority to enforce this law. This Act may be seen as a proactive move by the government to create a policing agency empowered to identify and restrict anticompetitive conduct which could not be foreseen by the US Congress or which did not properly fall under the scope of the Sherman or Clayton Acts. Here the Commission's activities range from investigating instances of deceptive ADVERTISING to promulgating trade regulation rules concerning products or industry activities.

Unlike the Antitrust Division of the Department of Justice, which is part of the executive branch of government, the FTC is a largely independent commission made up of five individuals who are appointed for seven year terms which are staggered to avoid coinciding with national political elections. By giving the Commission's trade regulation rules the full force of law, by enabling it to issue cease and desist orders, and by allowing it (in some instances) to require corrective action (e.g., for deceptive advertising), the Act yields considerable power to the FTC to enforce the law. The FTC maintains its own quasi judicial system, under which an administrative law judge may hand down an initial decision. Any such decision, though, is subject to appeal to the commissioners, and ultimately to the Federal courts. In addition, Congress determines the FTC's budget. This may also diminish the extent to which its decision making power is independent from political pressure.

As is the case in other antitrust statutes, interpretation of the language used is of central importance. Here, the words "unfair" and "deceptive" have been given great attention by the enforcers of the Act. Jorde, Lemley, and Mnookin (1996) demonstrate that the Act gives authority to the Commission to prohibit business conduct which *may* lead to a violation, and also empowers it with the authority to declare certain forms of conduct to be "unfair," even when that conduct violates neither the letter nor the spirit of other state or Federal antitrust laws. Many service industries, including automotive and TV repair, funeral services, and the optometry industry, as well as door to door sales activities have come under regulation by the FTC as a result of practices which have been identified as "unfair"; see, e.g., US Federal Trade Commission Staff Report (1974).

As Greer (1992) demonstrates, two standards have been established for the meaning of the term "deceptive." Over the past half century the courts, and for the most part the Commission itself, have required that an act need only to have the *capacity* or *tendency* to deceive a *substantial number* of buyers in some *material respect* to be declared a violation. Thus, the actual deception of a buyer, and deliberate intent to deceive, by the seller, need not be proven. During the 1980s the Commission altered its standard by requiring that the action under consideration be *likely* to deceive *reasonable* consumers and achieve *detrimental results* prior to its being declared to be in violation of the law. As Greer further points out, this new standard has not taken hold in the courts, nor has it been accepted by individual states, which also have statutes similar to the FTC Act.

See also *antitrust policy (US)*

Bibliography

Greer, D. F. (1992). *Industrial Organization and Public Policy*, 3rd edn. New York: Macmillan.

Howard, M. (1983). *Antitrust and Trade Regulation*. Englewood Cliffs, NJ: Prentice-Hall.

Jorde, T., Lemley, M., and Mnookin, R. (1996). *Gilbert Law Summaries: Antitrust*, 9th edn. Chicago: Harcourt Brace Legal and Professional Publications.

Martin, S. (1994). *Industrial Economics*, 2nd edn. New York: Macmillan.

US Federal Trade Commission Staff Report (1974). *Regulation of the Television Repair Industry in Louisiana and California*. Washington, DC: US Government Printing Office.

firm financial structure

John Edmunds and Roberto Bonifaz

Business firms acquire resources from lenders, from stockholders, and from other stakeholders including suppliers and employees. By convention, lenders are usually thought to include banks and bondholders, who lend money to the firm; stockholders are the owners of the capital which is permanently invested in the firm. Both are distinguished from trade creditors, employees, and other stakeholders who put resources into the firm in the form of merchandise or labor. So the resources the firm uses are classified in three broad categories according to the nature of the resources and terms on which they are placed: equity funds, debt, and inputs that are not in monetary form, e.g., merchandise and labor. The firm's financial structure is the term given to describe the specific composition of these inputs that a firm is using. For example, a firm may be financed in the following way:

Balance Sheet

Assets		Liabilities and Stockholders' Equity	
Liquid	10	Accounts Payable	20
Accounts Receivable	20	Accrued Wages	10
Inventory	20	Bank Loan Payable	50
Fixed Assets	50	Stockholders' Equity	20
Total	100	Total	100

This firm's financial structure is 50 percent debt, 20 percent equity, and 30 percent other liabilities. Some writers make a distinction between the terms financial structure and capital structure. Financial structure refers to all the resources the firm uses, and is expressed in terms of proportions of the entire amount.

Capital structure refers to the mix of resources that come from lenders and stockholders only (*see* CAPITAL MARKETS). Trade creditors, employees, and accrued taxes are not taken into account in discussing capital structure, because the resources that come from those sources arise in the routine course of business, and are not the result of financing activities that the firm arranges with its investment bankers to access the capital markets.

Each of the three groups (equity funds, debt, and inputs) has different incentives regarding how it would prefer the firm to be managed or how it should be handled in BANKRUPTCY proceedings. These differences in incentives may lead to differences in the management of the firm. For example, workers at a plant on the verge of bankruptcy would prefer that pension obligations receive first priority, which might put them at odds with the stockholders or trade creditors. On the other hand, trade creditors might prefer the immediate liquidation of the firm if they thought it likely they would receive full payment, but this short term focus might conflict with the interests of the firm's stockholders and workers (see Milgrom and Roberts, 1992).

Bibliography

Brigham, E. F. and Gapenski, L. C. (1997). *Intermediate Financial Management*, 5th edn. New York: International Thomson Publishing.

Milgrom, P. and Roberts, J. (1992). *Economics, Organization and Management*. Englewood Cliffs, NJ: Prentice-Hall.

first-mover advantages

Robert E. McAuliffe

A firm has first mover advantages when it can act before its rivals and achieve a relatively stronger competitive position. For example, the first firm to introduce a new product may have lower marketing costs because it has no competition or it may be able to accumulate production volume to lower costs along its LEARNING CURVE. Porter (1980) suggests that early ENTRY into an emerging industry can be an appropriate strategic choice when the early

firm can develop a reputation in the market, take advantage of the learning curve, or establish brand loyalty among consumers. At the same time, early entry can mean higher risks when the costs of establishing the market are high or when rapid technical advances may make the product obsolete.

Schmalensee (1982) provided a theoretical model showing that the first brand to satisfy consumers in a new market could establish the standard against which all other products were measured and this conferred lasting advantages to those brands. Glazer (1985) examined entry into the Iowa newspaper markets and found that first entrants into a successful market were more likely to survive than second entrants, but when all markets were considered (including those where the first entrant failed) the first firms were not more likely to survive than later entrants. Of course, early entry can be risky and Sutton (1991) argued that being first did not necessarily confer advantages, so markets must be evaluated on a case by case basis. Nevertheless, he found evidence of first mover advantages in the US and UK prepared soups industry, the European margarine market, and in the European soft drink market (*see* PIONEERING BRANDS).

Established firms in an industry have first mover advantages relative to potential entrants which they can use strategically. Existing firms may overinvest in capacity to deter entry, or place the entrant at a relative disadvantage, or they may pursue a PRODUCT PROLIFERATION strategy and crowd the product space (see Tirole, 1988). An existing monopolist also has a greater incentive to introduce new products or add to capacity in a growing market before an entrant appears since MONOPOLY profits will be higher than the duopoly profits a potential entrant might expect. This is known as a market preemption strategy (see Tirole, 1988).

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Industrial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Glazer, A. (1985). The advantages of being first. *American Economic Review*, **75**, 473 80.

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.

Schmalensee, R. (1982). Product differentiation advantages of pioneering brands. *American Economic Review*, **72**, 349 65.

Sutton, J. (1991). *Sunk Costs and Market Structure*. Cambridge, MA: MIT Press.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

fixed costs

Gilbert Becker

Fixed costs are defined to be those costs which are independent of the level of output in the SHORT RUN. As such, these costs often are incurred prior to the actual production of the good in question. Fixed costs include, but are not limited to, the following:

1 the cost of CAPITAL goods (e.g., buildings, machinery);
2 the cost of land;
3 property taxes; and
4 the cost of various types of insurance.

The total dollar expenditures for any or all of the above costs may change over time, but nonetheless do not vary with the number of units produced on a daily or monthly basis. As such, each of these costs is classified as fixed. Some payments to employees, notably workers or managers who receive a fixed salary which is not tied to production levels, are also fixed costs. In contrast, hourly wages, energy costs, and materials costs vary directly with the level of production and are classified as **variable costs**. The term **overhead costs** is often used by managers, but this term holds varying meanings for different managers who sometimes mix fixed and variable costs under this heading.

The importance to managers of the distinction between fixed costs and those which are variable is that the former costs should not be included in the decision as to the proper level of production that a firm should establish in the short run. The rule for PROFIT MAXIMIZATION requires that the firm equate its MARGINAL REVENUE and MARGINAL COST in order to find an ideal output. Since the firm's fixed costs do not influence the extra cost of

producing additional units of output, these fixed costs should not be considered in this decision. Of course, since fixed costs are part of the firm's TOTAL COSTS, they will help to determine the overall level of profits that the firm achieves.

The level of fixed costs may also play a role in determining ENTRY into and EXIT from an industry. Very high fixed costs may deter entry (*see* BARRIERS TO ENTRY). In addition some fixed costs, such as ADVERTISING costs, may not be recoverable by a firm which fails prior to the DEPRECIATION of its fixed assets. These fixed costs, classified as SUNK COSTS, may also influence entry and exit as they help to identify the level of risk facing the owners of the firm.

Bibliography

Maurice, S. and Thomas, C. (2002). *Managerial Economics*, 7th edn. Boston: McGraw-Hill Irwin.

Nicholson, W. (2000). *Intermediate Microeconomics and its Application*, 8th edn. New York: Harcourt.

Waldman, D. and Jensen, E. (2001). *Industrial Organization*, 2nd edn. Boston: Addison-Wesley.

fragmented industries

Robert E. McAuliffe

Porter (1980) defined a fragmented industry as one comprised of many small firms where the concentration ratio of the four largest firms was 40 percent or less. No firm in such an industry has a significant market share and so there are no "dominant players" that can exert a strong influence on the industry. There are several economic forces at work which can cause a MARKET STRUCTURE to be fragmented, and these are primarily cost conditions and demand.

An industry cannot be fragmented when there are significant ECONOMIES OF SCALE in production or BARRIERS TO ENTRY because these factors lead to market structures that are highly concentrated. Porter also identifies high inventory costs or variable sales as contributing factors, since they make it difficult to produce in high volumes to lower costs or invest in capacity which might reduce costs. High transportation costs will also tend to reduce the MINIMUM EFFICIENT SCALE plant and will create isolated geographic markets (*see* GEOGRAPHIC MARKET). Products and services which are highly customized for individual customers do not lend themselves to economies of scale, so the nature of the product plays a role in determining industry structure (see Caves and Williamson, 1985).

Consumer preferences on the demand side of the market may also cause an industry to be fragmented. If consumers prefer products which are customized or if they have very diverse tastes, then it will be difficult for any brand to acquire a sizable market share (see Tirole, 1988). Diverse consumer tastes are not sufficient for a fragmented industry, however, because firms could offer multiple brands (a product line) to increase sales and market share. Additional forces must be at work to cause a fragmented structure, such as DISECONOMIES OF SCALE or the absence of ECONOMIES OF SCOPE. Porter suggests that rapid product and style changes could have these effects and place larger firms at a disadvantage relative to smaller, quicker firms. Finally, an industry structure may be fragmented because the industry is in the early stages of its life cycle, and will become more concentrated as standards emerge (*see* INDUSTRY LIFE CYCLE).

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Industrial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Caves, R. E. and Williamson, P. J. (1985). What is product differentiation, really? *Journal of Industrial Economics*, **34**, 113 32.

Martin, S. (2001). *Advanced Industrial Economics*, 2nd edn. Oxford: Blackwell.

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

G

game theory

Sayed Ajaz Hussein

Many situations in economics involve *strategic interactions* between economic agents. In such situations, a decision by any one agent depends upon the decisions made by other agents. Game theory is the analysis of strategic interaction bet ween economic agents (players) and prediction of actions ("equilibrium") that players will take.

In COURNOT COMPETITION, for example, market price depends on the combined output produced by all firms in the market and firms must choose their output without knowing their rivals' outputs. Firms must, therefore, when choosing their own output, anticipate their rivals' choice of output, giving rise to a strategic interaction between firms. Game theory offers a framework to predict the actual choices made by firms. In contrast, a firm in PERFECT COMPE TITION will not anticipate its rivals' choice of output, since all firms are price takers. The choice of output is made by equating (own) MARGINAL COST with price, ignoring the choices made by other firms.

Here we discuss various types of games and the corresponding equilibrium concepts, focus ing on noncooperative games where players do not form agreements (or coalitions) with other players and instead compete with other players. For discussion of cooperative games see, e.g., Moulin (1988), Myerson (1991), or Osborne and Rubinstein (1994).

Games can be characterized by four criteria: static *or* dynamic, and games of complete infor mation *or* incomplete information. A game is static if it is played for a single period or where players do not observe other players move before making a move (known as *imperfect information*). A game is dynamic if played over many periods or where players observe other players' moves before making a move (known as *perfect infor mation*). Games of complete information are those in which each player has information about its rivals' precise characteristics (such as preferences, strategy spaces), while in games of incomplete information such information is missing. Corresponding to each type of game is an equilibrium concept summarized in the following table:

	Static	*Dynamic*
Complete information	Nash	Perfect Nash
Incomplete information	Bayesian Nash	Perfect Bayesian

In a *static game with complete information*, players do not observe others' moves before making their own. An example is a Cournot competition game for one period – firms must choose their output without knowing what other firms have chosen. Another example is the PRISONER'S DILEMMA game:

		Bill	
		Confess	Do not confess
Al	Confess	**-2, -2**	**0**, –3
	Do not confess	–3, **0**	–1, –1

In this case two prisoners, Bill and Al, are being interrogated about a crime. Each prisoner has two strategies to choose from: confess or do not confess. By convention, Al has the row strategies, while Bill has the column strategies. The entries in each cell indicate what each prisoner gets for any combination of strategies – the first number in any cell is Al's payoff, while the second number is Bill's payoff. For example, if Al confesses to the crime, while Bill does not, we arrive at the cell (0, −3): Al gets no jail time, while Bill gets three years. Predictions of actions in static games with complete information are made using the **Nash equilibrium** concept, defined as a set of strategies in which no player will deviate given the choices of other players.

The prisoner's dilemma game can be solved using the Nash equilibrium as follows: we look for a pair of strategies in which each is a best response to the other, akin to saying that no prisoner will deviate from his choice. The bold entries show the best response for Al to his opponent's moves. To see this, note that if Al confesses, the best response for Bill is to confess; if Al does not confess, the best response for Bill is again to confess. If Bill confesses, Al will want to confess; and if Bill does not confess, then Al will again want to confess. There is only one set of strategies (confess, confess) where no prisoner

will deviate given the other's choice – this is a Nash equilibrium.

The Nash equilibrium discussed here is one in pure strategies, so called because such a strategy is played with probability 1. But there may be situations in which players may use strategies with some probability between 0 and 1. Such strategies are known as *mixed strategies*. Nash's (1951) contribution was to show that for any game there will be at least one mixed strategy equilibrium (see also Fudenberg and Tirole, 1991).

The prisoner's dilemma game highlights the fact that a Nash equilibrium is a prediction about the outcome of a game; it is *not* a statement about the *optimality* of outcomes. Notice in the prisoner's dilemma game that (do not confess, do not confess) is a superior outcome to the Nash equilibrium (confess, confess), yet the players pursuing their self interest choose a suboptimal outcome. There are several ways to reach an optimal outcome, including changing the setting from a noncooperative game to a cooperative game and changing the game to a dynamic setting (see below).

The Entry game below shows how a Nash equilibrium need not be unique. Two stores are debating whether to enter a market or stay out:

	Stay out	Enter
Stay out	0, 0	**0, 4**
Enter	**4, 0**	–4, –4

The entries in the cells are profits. There are two Nash equilibria: (Stay out, Enter) and (Enter, Stay out). Beyond making a prediction, the concept of Nash equilibrium does not indicate which equilibrium will emerge: additional refinement criteria are needed to pick from multiple equilibria.

DYNAMIC GAMES WITH COMPLETE INFORMATION

These games are played over many periods. This implies that some players make their move *after* other players have made a move. An example is a Stackelberg duopoly game, in which one firm chooses its output, after which the second firm chooses its output. In such games, players who move first can influence the actions of players who move later (*see* FIRST MOVER ADVANTAGES). It emerges that the Nash equilibrium applied to dynamic games can give rise to absurd equilibria: the equilibrium may involve noncredible threats which players may not exercise when called upon to do so (*see* CREDIBLE STRATEGIES). To counteract this problem, dynamic games with complete information are solved using a perfect Nash equilibrium, which requires that strategies be in equilibrium not only on the equilibrium path, but *off* the equilibrium path as well. The strategies in a perfect Nash equilibrium form a Nash equilibrium in any subgame. For example, a firm ("entrant") is contemplating whether to enter a market ("in") or stay "out" (see Mas Colell, Whinston, and Green, 1995). The "incumbent" firm can choose to "fight" or "accommodate" the entrant (the following representation of games is known as the "extensive form"):
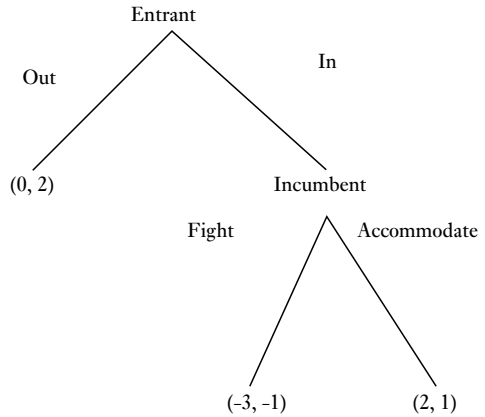
The first number represents the entrant's payoff and the second is the incumbent's payoff. To see the problem with the Nash equilibrium, the "normal" form of the box is useful:

		Incumbent	
		Fight if entrant plays "in"	Accommodate if entrant plays "in"
Entrant	Out	**0, 2**	0, 2
	In	–3, –1	**2, 1**

There are two Nash equilibria: (out, fight if entrant plays "in") and (in, accommodate if entrant plays "in"). The interpretation of the first equilibrium is a threat: the incumbent threatens to fight if the entrant enters the market. With this in mind, the entrant prefers to stay out. But this is not a credible threat: note that if the entrant actually enters, then the incumbent is better off accommodating rather than fighting: the equilibrium (out, fight if entrant plays "in") is not credible.

Perfect Nash equilibrium rules out noncredible equilibria since it requires that players' strategies are optimal even in off equilibrium paths. To locate the perfect Nash equilibrium, one can use backward induction to fold back the tree. We can locate the optimal response in the last subgame of the game (i.e., the subgame that follows "in"): the incumbent will accommodate if the entrant is in. The tree is then folded back to yield:
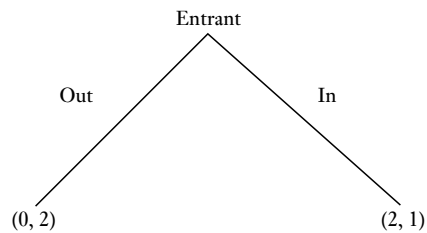
From this, we see the entrant, having considered the outcome if it enters the market, will choose to enter, and the incumbent will

accommodate. The noncredible equilibrium (out, fight if entrant plays "in") does not arise. The perfect Nash equilibrium is (in, accommo date if entrant plays "in").

Another interesting topic in dynamic games is that of the prisoner's dilemma (PD) repeated a finite number of times. In the static PD game, the Nash equilibrium is suboptimal. Using back ward induction, the finite repeated PD game again yields suboptimal results: at the final period, both players will confess. Stepping back, the penultimate period becomes the final period and again the players confess, and so on. If the games are not physically related (i.e., are independent), then the finite repeated game is a repetition of the static game.

This ceases to be the case when the game is played for infinite periods: other equilibria emerge, enabling the players to achieve optimal outcomes. For example, at any date, a player does not confess if and only if players have not confessed until that date, enabling players to sustain, under certain conditions, the optimal outcome. This result has useful implications for analysis of price wars, as it yields strategies for companies to get out of a price war using tit for tat strategies and other SIGNALING schemes (such as price matching).

STATIC GAMES WITH INCOMPLETE INFORMATION

Here, players don't know the history of the game and may be unaware of rivals' character istics. In a research and development (R&D) race, for example, firms may have to choose R&D expenditure without observing rivals' R&D ex penditures (static game of imperfect informa tion) and/or may lack knowledge about the efficacy of rivals' R&D efforts (incomplete in formation). This is another example of a Cour not OLIGOPOLY game where firms are unaware of rivals' cost functions.

Such games are solved using the Bayesian Nash equilibrium (Harsanyi, 1967–8). Players' characteristics are embodied in types (e.g., high efficacy of R&D efforts may be type 1 and low efficacy of R&D efforts may be type 2). All players are assumed to have common knowledge about the types and the statistical distribution from which these types are drawn. In the Bayesian equilibrium each player chooses his expected payoff contingent on his type and taking rivals' type contingent strategies as given.

The Bayesian Nash equilibrium is illustrated via a refined PD game (Mas Colell et al., 1995). Again, the prisoners can either confess to the crime or not confess to the crime. The first prisoner comes in one type while the second prisoner comes in two types (think of this as preferences depending on the day), with probability p and $(1-p)$. The tables below give the jail sentences for each type:

		Type 1 Prisoner 2 Probability p	
		Do not confess	Confess
Prisoner 1	Do not confess	0, –2	–10, –1
	Confess	–1, –10	–5, –5

		Type 2 Prisoner 2 Probability $1 - p$	
		Do not confess	Confess
Prisoner 1	Do not confess	0, –2	–10, –7
	Confess	–1, –10	–5, –11

To locate the Bayesian Nash equilibrium, note that prisoner 2 will always confess if he is type 1 and will always not confess if type 2. Prisoner 1 can then expect to get by not confessing: $p(-10)+(1-p)(0)$. By con

fessing, prisoner 1 can expect to get $p(-5)+(1-p)(-1)$. Thus, prisoner 1 will confess only if the EXPECTED VALUE from this exceeds the expected value from not confessing: $p(-5)+(1-p)(-1) > p(-1)+$

$(1 - p)(0)$, implying that if $p > 1/6$ (the prob ability that prisoner 2 is of type 1), prisoner 1 should confess to the crime, if $p < 1/6$ pris oner 1 should not confess, and if $p = 1/6$ he should be indifferent.

DYNAMIC GAMES WITH INCOMPLETE INFORMATION

In these examples, the equilibrium concept per fect Bayesian equilibrium (PBE) combines the Bayesian equilibrium (for incomplete informa tion games) with perfect equilibrium (for dy namic games). In such games, whenever a player makes a move, it divulges some informa tion about its type to other players who update their beliefs (using a Bayesian updating process) about other players. A PBE thus requires that strategies are optimal given beliefs, and that beliefs are obtained from strategies and observed actions using Bayes's rule. PBE can be found in games of reputation, warranty as a signal of quality, and sequential bargaining.

As can be seen from these examples, game theory offers numerous formats and solutions for games depending on the type of game being played. Future research will provide even more insight into games and their economic outcomes.

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Indus trial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Fudenberg, D. and Tirole, J. (1991). *Game Theory*. Cambridge, MA: MIT Press.

Harsanyi, J. (1967 8). Games with incomplete informa tion played by Bayesian players. *Management Science*, **14**, 159 82, 320 34, 486 502.

Mas-Colell, A., Whinston, M. D., and Green, J. (1995). *Microeconomic Theory*. New York: Oxford University Press.

Moulin, H. (1988). *Axioms of Cooperative Game Theory*. New York: Cambridge University Press.

Myerson, R. (1991). *Game Theory: Analysis of Conflict*. New York: Cambridge University Press.

Nash, J. F. (1951). Non-cooperative games. *Annals of Mathematics*, **54**, 289 95.

Osborne, M. J. and Rubinstein, A. (1994). *A Course in Game Theory*. Cambridge, MA: MIT Press.

geographic market

Gilbert Becker

One of the two boundaries which establish the scope of a market in an economic sense is the relevant geographic area of the market in ques tion. Since all markets are not national in scope, it is often difficult to properly define the geo graphic area in which a group of rivals compete. This area, ideally defined, must be broad enough to encompass all rivals which are reasonably seen as being in competition with one another while not being so broad as to include firms that are in some sense not actual or potential competitors. Moreover, it should include all sources of supply of this product which could or do impact market outcomes, especially price.

The importance of accurately defining the geographic market lies primarily in the applica tion of antitrust policy (*see* ANTITRUST POLICY (US)). For example, Section 2 of the SHERMAN ACT deals with attempts to monopolize a market, while Section 7 of the CLAYTON ACT places restrictions against mergers which lessen competition in any section of the country. The structure of a market and its degree of competi tion depend, in part, on the number of rivals competing with one another (*see* COMPETI TION; MARKET STRUCTURE). Hence, the de termination of whether an antitrust law violation has occurred depends in part on a proper count of the number of rivals existing in the market in question, and this requires the delineation of an appropriate geographic area to be considered.

While legal boundaries, such as interstate dif ferences in commercial practice laws, or topo graphical boundaries, such as mountains or rivers, often guide the courts in their determin ation of a relevant geographic market, economic analysis has also been of assistance in this effort. Elzinga and Hogarty (1973) developed one of the first tests, known as the "20 percent rule," which remains to date an important practical guide. The rule holds that a reasonable boundary has been established if less than 20 percent of the product that is consumed within the area under consideration is imported from outside of that area, and less than 20 percent of the product that is produced within this same area is shipped

outside of that area for consumption in a differ ent region.

The logic here is that when the conditions of the rule hold, the great majority of the firms affecting the market supply and demand, and therefore price, have been identified. If the actual figure in the first instance exceeds 20 percent, then rivals whose output may signifi cantly impact the market price are erroneously being ignored and the geographic boundary should be expanded to include them. If the second part of the rule is being violated and a significant portion of the supply of the product is being shipped beyond the geographic area being considered, this offers another indication that suppliers are not facing transportation and dis tribution costs sufficiently great so as to be con sidered prohibitive. Here again the geographic boundary under consideration should be expanded. Several other theoretical and practical measures of market boundaries exist (*see* MARKET DEFINITION).

See also *antitrust policy (US); Clayton Act; merger guidelines, 1992–7; Sherman Act*

Bibliography

Boyer, K. (1979). Industry boundaries. In T. Calvani and J. Siegfried (eds.), *Economic Analysis and Antitrust Law*. Boston: Little, Brown.

Carlton, D. W. and Perloff, J. M. (2000). *Modern Indus trial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Elzinga, K. and Hogarty, T. (1973). The problem of geographic market delineation in anti-merger suits. *Antitrust Bulletin*, **45**, 45 81.

Shepherd, W. and Shepherd, J. (2004). *The Economics of Industrial Organization*, 5th edn. Long Grove: Wave-land Press.

gross domestic product

Robert E. McAuliffe

Gross domestic product (GDP) is a measure of a country's output and is the current dollar value of all final goods and services produced for the market by those resources within the country's borders per period of time. GDP does not in clude the DEPRECIATION of capital equipment used to produce output over the period. Since GDP is measured in current dollars, it may rise either because current output increased or be cause the general level of prices increased (IN FLATION occurred). To compensate for the effects of changes in the purchasing power of the domestic currency, a measure of general price change is used to translate current dollars into constant purchasing power (or base year) dollars. When current dollar figures (nominal values) are adjusted by the appropriate price index (such as the GDP deflator or the consumer price index), the result is an inflation adjusted figure (i.e., a real value). Thus when real GDP has increased, a nation has produced more output.

For example, nominal GDP in the US grew from $10.542 trillion in the third quarter of 2002 to $11.107 trillion in the third quarter of 2003. But after adjusting for changes in the price level, real GDP only rose from 10.1284 trillion 2000 dollars to 10.4938 trillion 2000 dollars, reflecting the economic growth that occurred during that period. Managers in industries whose products follow the business cycle, such as automobiles, housing construction, and durable goods, can use GDP figures to forecast business conditions and sales.

See also *nominal income and prices*

Bibliography

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Sommers, A. T. (1985). *The US Economy Demystified*. Lexington, MA: D. C. Heath.

H

Herfindahl-Hirschman index

Roger Tutterow

While a complete analysis of market structure may require information on the market share of each competitor, industrial concentration is fre quently measured with a single summary index. Such indices are desirable for both econometric research and in establishing antitrust guidelines. The Herfindahl Hirschman index (HHI) is among the most widely accepted measures of market concentration. The index is defined as:

$$\text{HHI} = \sum_{i\ 1}^{n} S_i^2$$

where S_i denotes the market share of the ith firm in an industry composed of n firms. The HHI ranges between negligible values for perfectly competitive industries (*see* PERFECT COMPE TITION) and a maximum value of 1.0 for a pure MONOPOLY (or 10,000 if market shares are measured as a percentage).

Considerable theoretical work has related the HHI to industrial behavior. Stigler (1964) rela ted the HHI to the pricing behavior of oligop olies. Adelman (1969) suggested that the HHI has an intuitive "numbers equivalent" interpret ation in that 1/HHI is the number of equal sized firms that will produce a given HHI. Cowling and Waterson (1976) demonstrated that the HHI may be related to profitability in industries with constant MARGINAL COST.

Since the 1980s, the Department of Justice has utilized the HHI as a primary measure of market concentration. For the calculations used in their merger guidelines, the decimal point in the firms' market shares is ignored, thus produ cing an HHI which is larger by a multiple of 10,000. Accepting this modification, current guidelines specify that markets in which the HHI exceeds 1,800 are said to be highly concen trated. Thus, mergers resulting in an HHI above 1,800 are likely to attract in antitrust litigation. Mergers producing an HHI between 1,000 and 1,800 are likely challenged only if the HHI in creases in excess of 100 units. Mergers resulting in an HHI of less than 1,000 are unlikely to meet objection (*see* ANTITRUST POLICY (US); MERGER GUIDELINES, 1992–7).

Bibliography

Adelman, M. (1969). Comment on the "H" concentration measure as a numbers equivalent. *Review of Economics and Statistics*, **51**, 99 101.

Carlton, D. W. and Perloff, J. M. (2000). *Modern Indus trial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Cowling, K. and Waterson, M. (1976). Price cost margins and market structure. *Economica*, **43**, 267 74.

Martin, S. (2001). *Advanced Industrial Economics*, 2nd edn. Oxford: Blackwell.

Stigler, G. (1964). The theory of oligopoly. *Journal of Political Economy*, **72**, 44 61.

heteroskedasticity

Alastair McFarlane

Heteroskedasticity is present when the VARI ANCE of the error terms in a LINEAR REGRES SION is not identical for all observations. This violation of the homoskedastic assumption underlying the classical regression model is common in CROSS SECTION ANALYSIS of microeconomic data where the variance of the explained variable is often correlated with the size of an explanatory variable. An example is the consumption function of a family where there is

a greater variation in expenditures as family income rises.

Under heteroskedasticity, the ordinary least squares estimator is not efficient, that is, there exist other estimators that possess a lower variance. Another implication of heteroskedasticity is that although the ordinary least squares estimator of the regression coefficient will be unbiased, the estimator of the STANDARD ERROR OF THE COEFFICIENT will be biased. This bias causes hypothesis tests to be invalid.

One method of detecting heteroskedasticity is to examine the residuals plotted against particular explanatory variables. A more formal approach is the Glesjer test, which involves regressing the absolute value of the residual on a selected independent variable (or some function of that variable). Other frequently used tests are the White test, the Goldfeld Quandt test, and the Breusch Pagan test (see Maddala, 2001).

One solution to the problem of heteroskedasticity is to specify the variables in a manner that reduces the error variance. Common practices are to deflate all variables by some measure of size or to transform the data to logs. A second approach is to use a regression technique known as weighted least squares where observations with smaller variances receive a larger weight in the computation of the regression coefficients.

Bibliography

Greene, W. H. (2002). *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.

Gujarati, D. N. (2002). *Basic Econometrics*, 4th edn. New York: McGraw-Hill Irwin.

Kennedy, P. (2003). *A Guide to Econometrics*, 5th edn. Cambridge, MA: MIT Press.

Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.

hierarchy

Robert E. McAuliffe

As Coase (1937) observed, markets and firms are alternative methods of organizing production and allocating resources. Given the prevalence of large business organizations in market economies, it is important to analyze how corporations operate and make decisions. The term hierarchy refers to the hierarchical structure of business organizations, and Williamson (1975) examined the factors that affect the decision to produce inputs internally within the firm rather than purchase them externally from the market. Coase (1937) argued that it was costly for firms to use the market because of TRANSACTIONS COSTS. Therefore firms would expand the scope of their operations to reduce these costs. But as the firm increases in size, the costs of coordinating decisions within the firm increase. The optimal size of the firm is the size that minimizes the total transactions costs of using the market and internal production.

Williamson (1985) stressed that transactions costs are affected by UNCERTAINTY, the frequency of the transaction, and the specificity of assets in which the parties must invest (*see* ASSET SPECIFICITY). When a firm contracts in the market and requires a supplier to purchase specific equipment which has few alternative uses, the supplier could be vulnerable if the purchasing firm decides to purchase its inputs elsewhere. Detailed, long term CONTRACTS could be written to reduce the supplier's risk, but all future contingencies cannot be specified and so uncertainty and INCOMPLETE CONTRACTS may prevent the transaction from occurring. In such a case, VERTICAL INTEGRATION may solve the problem by internalizing the transaction. The governance of the market is then replaced by the hierarchical governance of the firm.

Hierarchies perform a variety of functions in the market. For example, a hierarchy can be viewed as an information system, since the structure of an organization affects the flow of information within the firm. As Radner (1992) suggests, economic EFFICIENCY requires specialization, but specialization requires decentralized decision making which introduces additional coordination problems within the firm. Another interesting feature of hierarchies is the internal labor market they create. It is important for any economic organization to properly match labor skills with job requirements. Holmstrom and Tirole (1989) suggest that when there is uncertainty about the job characteristics or the worker's skills, the hierarchy can act as a filter providing information about workers' abilities and matching them with

the appropriate current position and providing a learning environment for future positions. In fact, Holmstrom and Tirole suggest that when a mismatch between the worker and the job is too costly, long career paths with assignments throughout the hierarchy provide more infor mation about the worker's abilities (as with doctors and airline pilots).

Bibliography

Coase, R. H. (1937). The nature of the firm. *Economica*, **4**, 386 405.

Holmstrom, B. R. and Tirole, J. (1989). The theory of the firm. In R. Schmalensee and R. D. Willig (eds.), *Hand book of Industrial Organization*. New York: North-Hol land.

Radner, R. (1992). Hierarchy: The economics of man aging. *Journal of Economic Literature*, **30**, 1382 415.

Shughart, W. F. (1990). *The Organization of Industry*. Boston: Richard D. Irwin.

Williamson, O. E. (1975). *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: Free Press.

Williamson, O. E. (1985). *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. New York: Free Press.

horizontal merger guidelines

Gilbert Becker

Section 7 of the US CLAYTON ACT proscribes mergers which may substantially lessen COM PETITION. The language of the Act leaves open to interpretation, by the courts and the antitrust enforcement agencies, the exact nature of the circumstances under which there exists a viola tion, and varying interpretations have been offered (*see* ANTITRUST POLICY (US)). On three occasions since the late 1960s the federal government, in an effort to clarify the antitrust enforcement philosophy of the existing adminis tration, has issued guidelines indicating the con ditions under which the challenge of a merger will be likely.

POLICY DEVELOPMENT

In 1968 the Department of Justice (DOJ) pre sented the first set of guidelines. These guide lines followed a strict *structuralist* approach (*see* STRUCTURE CONDUCT PERFORMANCE PARADIGM) whereby markets were classified as "highly concentrated" or "less highly concen trated" and the decision to challenge a merger between rival firms was essentially based on one criterion, that being whether the market shares of the firms in question were unacceptably high. Specific rules were established by the guidelines. For example, in a highly concentrated industry, one in which the four firm concentration ratio (CR_4) was above 75 percent (*see* CONCENTRA TION INDICES), a merger between an acquiring firm holding a market share in excess of 10 percent and a firm to be acquired having a market share in excess of 2 percent was to be challenged (see US DOJ, 1982, for a complete statement of these guidelines and those developed in 1982). These guidelines offered a possible exemption in the case where the acqui red firm was failing and no other more suitable purchaser was forthcoming, but little else by way of a defense was to be allowed for mergers not falling within these rules. Moreover, the guide lines indicated willingness by the government to challenge an acquisition, even when the market shares were acceptably low, if the firm to be acquired had shown to be an aggressive, com petitive force, or if the merger was part of a trend toward increasing concentration in a market.

The revised 1982–4 guidelines and those which are currently in force (*see* MERGER GUIDELINES, 1992–7) present a substantial change from the first set of guidelines in terms of both the level of economic analysis generated and their economic philosophy. The prevailing criticism of the first set of guidelines is that they were too simplistic and too rigid, and that the long evidentiary record of the link between MARKET STRUCTURE and MARKET PERFORMANCE was weak and not sufficiently compelling to justify the strict rules that had been used. Thus, while both of the new sets of guidelines maintain an analysis of market structure as part of their approach, they replace the rigid rules approach with a case by case approach that uses additional economic factors, beyond market shares, in order to help indicate the *probability* of a challenge.

The new guidelines since 1982 replace the CR_4 with the HERFINDAHL HIRSCHMAN INDEX, HHI, as the preferred measure of indus try concentration. In addition, three levels of

concentration are identified and used in the first step of the evaluation process:

1 low concentration (where the post merger HHI < 1,000), in which a challenge is *un likely*;
2 moderate concentration (where after the merger 1,000 < HHI < 1,800), in which a challenge is *more likely than not* if the pro posed merger increases the HHI by more than 100 points; and
3 high concentration (where the post merger HHI > 1,800), in which a challenge is *likely* if the HHI increases by more than 50 points.

From there, the new guidelines significantly expand the use of economic analysis relative to their earlier counterparts. For example, a princi pal concern of antitrust enforcers dealing with horizontal mergers is the possibility that the declining number of rivals may facilitate the abuse of MARKET POWER through COLLU SION. As a result, the new guidelines investigate other factors which are known to influence the possibility of collusion. Since economic theory indicates that collusion is easier to achieve in a market where the good in question is homoge neous and where information on rivals' prices and sales is readily available to competitors, these factors will also be considered in evaluating the merger. Once all such factors are considered, if the evidence indicates that the market in ques tion is one in which collusion is easy, then the merger in question will be *more likely* to be challenged, all else being equal, than one involv ing a market where collusion is seen as being difficult.

Similarly, the guidelines detail a number of factors which are to be used in defining markets and present a new economic approach to defin ing markets based on firms' ability to introduce and maintain a SSNIP – a small but significant nontransitory increase in prices – to a market (*see* MARKET DEFINITION). Moreover, the new guidelines call for increased emphasis being placed on the condition of ENTRY into the market, under the theory that potential entrants may be able to discipline firms that attempt to abuse market power resulting from a merger (*see* CONTESTABLE MARKETS).

Finally, the latest two sets of guidelines offer significant changes concerning the question of EFFICIENCY. It has long been recognized that mergers may benefit consumers, even if compe tition is reduced, by allowing ECONOMIES OF SCALE and other cost saving efficiencies to take place. Williamson (1968) developed a framework for analyzing the merits of a horizontal merger in which such efficiencies occur. Using a simple cost benefit approach (*see* PROFIT MAXIMIZA TION), he argued that a merger should be appro ved if the potential gains to society from the lower costs of production exceed the potential cost to society from the DEADWEIGHT LOSS resulting from increased market power. More over, his model indicates that a relatively small gain in efficiency was necessary to offset the social loss and, as a result, the efficiency resulting from a merger should be considered as an important factor. Although neither the earlier (1968) guidelines nor the courts nor the Clayton Act itself has accepted efficiency con siderations as a defense of a merger, the newer guidelines treat efficiency much more favorably than in the past.

The 1992 guidelines, following the letter and spirit of those developed a decade earlier, offer some additional important refinements to the economic analysis of proposed mergers. First, the new guidelines more forcefully indicate that a merger can be challenged because of an antici pated transfer of wealth from consumers to pro ducers – i.e., ECONOMIC PROFIT. While the 1982 guidelines had focused primarily on the misallocation of resources, the current guidelines clearly offer both issues as a reason for govern mental action against a proposed merger. This, in theory, may substantially tip the cost benefit scales used in Williamson's approach. To the extent that the current and future administra tions measure and include excess profits as a cost of a proposed merger, this will, as suggested in Meehan (1977), dramatically increase the effi ciency gains needed to justify most mergers. In addition, this would bring US policy closer to that of the European Union (EU), which exam ines *abuse of power* and, in some settings, requires that consumers are allowed a *fair share of the benefits* achieved from technical efficiency gains (*see* EU COMPETITION POLICY, 2004). More over, the 1992 guidelines, while recognizing both

efficiency and entry conditions as being import ant, require a higher standard of proof than their earlier counterparts. This became increasingly apparent in 1997, when the government made minor revisions to the standards set in 1992.

Another change in the 1992 guidelines is that the language dealing with the likelihood of a challenge is, by the government's own assess ment, weaker than in the past. This is particu larly true in the moderately and highly concentrated industry categories. For example, in the latter category, the 1984 guidelines indi cated that a challenge would occur "except in extraordinary cases." The current guidelines, while presuming that anticompetitive effects arise from such increases in concentration, indi cate that demonstrating other importance effects of the merger may overcome the initial presump tion (see US DOJ and Federal Trade Commis sion, 1992: 1–7). As a result, although during the 1980s the US experienced its second largest wave of merger activity of the twentieth century, this movement was exceeded in size by that which occurred during the 1990s.

Policy Debate

One of the areas of greatest debate concerning antitrust policy has been that of the merger guidelines. The first of many issues debated is that of the merits of a more strict "rules" ap proach as offered in the first guidelines, in com parison with the more open "case" approach used during the last two decades. Meehan (1977), in a survey of the early literature defending the rules approach, cites several argu ments. The first is that it is a much less costly method since the case approach requires gathering a great deal of information, much of which comes from the firms in question and therefore must be verified due to the incentives of these firms to provide biased information. Second, the difficulty in the case approach of evaluating the social loss is recognized as a prob lem.

Third, and perhaps most importantly, signifi cant problems arise using Williamson's model to defend the efficiency argument. Meehan cites two studies which find that economies are often not a primary goal cited by managers involved in mergers and, more importantly, that in nearly half of the cases of mergers examined, no attempt was made to achieve potential econ omies (see synergies). Moreover, evidence is offered that the gain in market power accrued from these mergers was a primary cause for the diminished desire to achieve the efficiencies that were available. In addition, although mergers may offer cost savings through multiplant oper ations or through the rationalization (removal) of excess capacity, the denial of such mergers may only delay these cost savings as competition will induce the individual firms to expand or contract on their own. Finally, the argument has been made that since mergers may increase economic profit, firms may misallocate (from society's point of view) a significant amount of resources in an effort to achieve these profits, and that this misallocation should be included in any cost benefit analysis.

Meehan's solution to the efficiency defense question called for a modification of the 1968 guidelines to ignore mergers in which CR_4 is less than 50 percent, as the threat of market power is minimal. This would also allow for significant economies to be achieved where they are most likely to occur, i.e., in mergers among small and medium sized firms (see minimum efficient scale). The three levels of HHI used in current merger policy offer a movement in this direc tion.

In evaluating merger policy following the introduction of the 1982–4 guidelines, several scholars have offered policy solutions. Fisher (1987) argued for a very high standard of proof for firms claiming efficiencies and a refusal of the claims if an alternative method is available for achieving the economies. Schmalensee (1987) argued that efficiency should only be used as a tiebreaker of last resort, and should not save an otherwise objectionable merger. In general, these and other scholars favor the new case ap proach in its treatment of market definition and entry, while disagreement arises as to the rewrit ing of the Clayton Act to specifically allow an efficiency defense. Miller (1984) takes the strong stance that the 1984 revision of the 1982–4 guidelines, by evolving to a cost benefit ap proach wherein efficiency is considered, "borders on executive legislation" as courts are unlikely to be able to test this doctrine since the government is not bringing cases where efficien cies are great.

During the past two decades, critics have on numerous occasions cited lax antitrust policy as being a major factor contributing to the level of merger activity over this period. In one recent sharp critique of both the 1982–4 and 1992 guidelines, Shepherd, Shepherd, and Shepherd (2000) argue that they have lost the focus of their original intent, which was to clarify the government's position on Section 7 of the Clayton Act and to foster effective competition and all of its benefits. The 1968 guidelines were clear and relatively simple rules for the business community to understand, and were consistent with court interpretations of that time concerning borderline merger cases. The new guidelines, from Shepherd et al.'s point of view, delineate much of the current economic knowledge concerning markets and develop numerous tests and screens in the process of evaluating a merger, but in doing so no longer reduce the uncertainty on the part of businesses in the US as to which mergers will be challenged. Blumenthal (1993), while favoring the approach of the new guidelines, reinforces the problem of their lack of clarity by identifying 31 new issues that are raised by these guidelines (see *Antitrust Bulletin*, 1993, for a symposium on the 1992 guidelines).

Moreover, Shepherd et al. argue that the proliferation of the factors identified in the new guidelines deliberately furthers a bias, held by the government in the past two decades, toward favoring big business. The potential time and cost to small firms of a court challenge would be prohibitive, while large firms wishing to merge could impose significant costs on the government agency opposing the merger by forcing them to develop the extensive evidence required by the new guidelines. Shepherd et al. also argue that the bias toward bigness is enhanced by the entry criteria in the guidelines. The condition of entry is seen as overshadowing the importance of the number and size of actual competitors, leading to mergers having been allowed wherein the new firm has held anywhere from 25 percent to 60 percent of the market output. They argue that the DOJ guidelines tend to expand beyond those laid down by the courts, and as such the focus of the new guidelines is inconsistent both with the application of other US antitrust laws and with recent European antitrust decisions. As a result, they call for a simpler set of rules, focused on single firm dominance, wherein no firm is allowed to achieve a market share of greater than 25 percent via merger.

In sum, the horizontal merger guidelines of the past 30 years have continued to develop in terms of their economic content and analysis. As sharp differences in economic philosophies exist, these guidelines continue to be a source of contention among antitrust policy makers, scholars, and industry leaders.

Bibliography

Blumenthal, W. (1993). Thirty-one merger policy questions still lingering after the 1992 guidelines. In D. Scheffman (guest ed.), *Antitrust Bulletin*, **38**, symposium on new 1992 merger guidelines.

Church, J. and Ware, R. (2000). *Industrial Organization: A Strategic Approach*. Boston: McGraw-Hill Irwin.

Fisher, F. (1987). Horizontal mergers: Triage and treatment. *Economic Perspectives*, **1**, 23 40.

Meehan, J. (1977). Rules vs. discretion: A reevaluation of the merger guidelines for horizontal mergers. *Antitrust Bulletin*, **20**, 769 95.

Miller, R. (1984). Notes on the 1984 merger guidelines: Clarification of the policy or repeal of the Cellar-Kefauver Act? *Antitrust Bulletin*, **29**, 653 62.

Schmalensee, R. (1987). Horizontal merger policy: Problems and changes. *Economic Perspectives*, **1**, 41 54.

Shepherd, G., Shepherd, H., and Shepherd, W. (2000). Sharper focus: Market shares in the merger guidelines. *Antitrust Bulletin*, **45**, 835 85.

US Department of Justice (1982). Merger guidelines, 1966 and 1982. *Antitrust Bulletin*, **27**, 633 83.

US Department of Justice and Federal Trade Commission (April 2, 1992; revised April 8, 1997). *Horizontal Merger Guidelines*. Washington, DC: US Government Printing Office.

US Federal Trade Commission Bureau of Competition (1992). *How 1992 Guidelines Differ from Prior Agency Standards*. Washington, DC: US Federal Trade Commission.

Williamson, O. (1968). Economies as an antitrust defense: The welfare tradeoffs. *American Economic Review*, **58**, 18 36.

identification problem

Robert E. McAuliffe

Two econometric issues arise whenever a re searcher attempts to estimate simultaneous relationships: the identification problem and SIMULTANEOUS EQUATIONS BIAS. When an equation is not identified, the researcher cannot estimate the underlying equation (such as the demand or supply curve). When an equation is part of a simultaneous system of equations, such as the supply and demand for a product, direct estimates obtained from a LINEAR REGRES SION will be biased and other estimation methods must be employed.

To understand the basic nature of the identi fication problem, consider a competitive market where the prices and quantities observed result from the interaction of *both* supply and demand. Given that both supply and demand determine observed data for prices and quantities, how can a researcher know that the resulting EQUILIB RIUM data points reflect the underlying demand curve or supply curve? This is the issue of iden tification. Without any additional information beyond observed data on prices and quantities, a researcher cannot know whether the observed points reveal a demand or supply relation, or a weighted average of both. However, if additional information is available, it may be possible to identify either the DEMAND FUNCTION, the supply curve, or both.

Consider the system of supply and DEMAND CURVES for any competitive market below:

$$Q_t^d = \alpha_1 + \alpha_2 P + \varepsilon_t^d \qquad (1)$$

$$Q_t^s = \beta_1 + \beta_2 P + \varepsilon_t^s \qquad (2)$$

$$Q_t^d = Q_t^s \qquad (3)$$

where Q is the quantity supplied or demanded, P is the price of the product, and the ε_t^i are random disturbances to demand and supply. Equations (1)–(3) are referred to as *structural equations* which reflect specific features of the market. These equations are mutually dependent since price and quantity cannot be determined with out all of them. Even when a researcher is only interested in estimating one of these equations (such as demand), if price and quantity are deter mined by both supply and demand, the demand curve must be regarded as part of a system of equations (see Greene, 2002; Maddala, 2001).

As written, neither demand nor supply is identified above without additional information, since each curve contains identical information relating quantity to price. Because both equa tions are not identified, their underlying param eters (the αs and βs) cannot be estimated from the data. However, if a researcher knew that there was no disturbance to demand (or very little random variation in demand), this add itional information would serve to identify the demand curve. The reason is that in this case the researcher would know that demand was given by:

$$Q_t^d = \alpha_1 + \alpha_2 P \qquad (1')$$

which means that the observed fluctuations in price and output in the market were due to variations in supply. That is, if the demand curve had little or no variation, then the re searcher would know that observed movements in prices and quantities were caused by shifts in supply along a (relatively) constant demand curve. The resulting data on prices and quan tities would thus trace out the demand curve as shown in figure 1.

Price
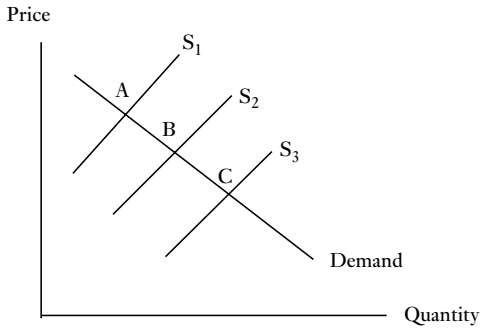
Figure 1   Shifts in supply identify the demand curve

This example illustrates the identification problem and how it may be solved. To state the requirement loosely at the moment, for a given equation to be identified, that equation must *exclude* variables which are included in other equations in the model or other restrictions must be imposed on the data. In the example above, the demand curve (1′) would be identified but the supply curve is not identified because the disturbance term is excluded from the demand curve, but there are no variables excluded from the supply curve which are included in demand. It should also be noted that identification is a property of the data available: a researcher could not simply assume that there was no variation in demand to identify the demand curve if it were not the case because the resulting data would reflect shifts in both demand and supply.

Consider the system of supply and demand equations below:

$$Q_t^d = \alpha_1 + \alpha_2 P + \alpha_3 I + \varepsilon_t^d \qquad (1a)$$

$$Q_t^s = \beta_1 + \beta_2 P + \beta_3 C + \varepsilon_t^s \qquad (2a)$$

$$Q_t^d = Q_t^s \qquad (3)$$

where I is consumer income (an exogenous variable) and C is the cost of resources in production (also an exogenous variable in this system). In this case, both supply and demand are identified. The demand curve is identified because resource costs which shift supply are excluded from the demand curve. Supply is identified because variations in consumer income will shift demand without changing supply. As long as

the parameters α_3 and β_3 are not equal to zero, then changes in income and costs will serve to identify supply and demand, respectively.

More generally, a single equation in a system of simultaneous equations may be *exactly identified*, *over identified*, or *under identified* depending (in part) on whether the exclusion restrictions necessary to identify the equation are met. In the modified example above, the demand curve (1a) is exactly identified, as is the supply curve (2a). To be identified, any given equation must meet the *order* and *rank* conditions described below (see Greene, 2002; Maddala, 2001).

Consider a system of equations with N *endogenous* variables (that is, variables which are to be explained by the model such as price and quantity in the examples above) and K *exogenous* or *predetermined* variables. Exogenous variables are those which are not explained by the model and which are not correlated with the disturbance terms in each equation, such as income and resource costs in the example above (*see* ERROR TERMS IN REGRESSION), while predetermined variables are past values of the endogenous variables. For linear systems the order condition requires that the total number of endogenous, exogenous, and predetermined variables *excluded* from the equation must equal the number of endogenous variables *included* in that equation minus 1. So if there are J endogenous, exogenous, and predetermined variables excluded from an equation in a system of N endogenous variables, then:

1   The equation is exactly identified when $J = N - 1$.
2   The equation is over identified when $J > N - 1$.
3   The equation is under identified when $J < N - 1$.

When an equation is under identified, its parameters cannot be estimated. The order condition is necessary for identification but is not sufficient. In most cases, the rank condition for identification will be satisfied when the order condition is satisfied, but exceptions can occur.

The rank condition for identification is both necessary and sufficient for identification and refers to the rank of largest number of linearly independent rows or columns of a matrix (the

system of equations). Maddala (2001) provides a convenient summary of the requirements for identification: consider a system of three equa tions with three endogenous variables, y_1, y_2, y_3, and three exogenous variables, x_1, x_2, x_3 (this method easily generalizes to systems with different numbers of equations and variables than those used in this example). Consider the rows and columns for the equations as if in a spreadsheet, and place an X when a variable appears in the equation and 0 when it does not appear. For a system such as the one below, we can check each equation to see if it is identified.

Equation	y_1	y_2	y_3	x_1	x_2	x_3
1	X	0	0	0	0	X
2	X	0	X	0	0	X
3	0	X	X	X	X	0

The rules Maddala provides for identifying any equation are:

1 Delete the row of the equation being con sidered.
2 Take the columns with the 0 entries for that row.
3 Given these remaining columns, there must be $(N-1)$ rows and columns which do not have only zeros, and no row (or column) can be proportional to any other row (or column) for all parameter values. If these require ments are fulfilled, then the equation is iden tified.

This version of the rank condition is both ne cessary and sufficient for an equation to be identified. Consider the table above. There are three endogenous variables $(N = 3)$, so the order condition requires that at least two vari ables are excluded from each equation. The first equation has four variables missing and is over identified, the second equation has three variables excluded and is also over identified, while the third equation has two variables ex cluded and is exactly identified. Therefore all three equations meet the order condition for identification.

For the rank condition we require that there be $N - 1 = 2$ nonzero rows and columns. For equation 1, delete that row and take those columns where zeros appear in the equation. The columns with zeros for equation (1) are:

$$
\begin{array}{cccc}
0 & X & 0 & 0 \\
X & X & X & X
\end{array}
$$

Since we have two rows (and two columns) which are not completely filled with zeros, the equation is identified. To apply the rank condi tion to equation (2), delete that row and take those columns where zeros appear to obtain:

$$
\begin{array}{ccc}
0 & 0 & 0 \\
X & X & X
\end{array}
$$

Since an entire row contains zeros and only one row does not, equation (2) does not meet the rank condition of having $(N-1)$ rows and columns that are not all zeros. Thus equation (2) is not identified despite the fact that the order condition was fulfilled and so its parameters cannot be estimated. As the reader can verify, the third equation also meets the rank condition for identification. The identification problem can also be expressed in terms of the relationship between the structural equations of the model and the *reduced form equations* (*see* ESTIMATING DEMAND). For more advanced treatments, see Fisher (1966) and Hsiao (1983).

Bibliography

Fisher, F. M. (1966). *The Identification Problem in Econo metrics*. New York: McGraw-Hill.
Greene, W. H. (2002). *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.
Hsiao, C. (1983). Identification. In Z. Griliches and M. D. Intriligator (eds.), *The Handbook of Econometrics*. Amsterdam: North-Holland.
Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.

imperfect information

Alexandra Bernasek

Perfect information is the strongest information requirement that can be assumed in an economic

model. Traditionally, it has been a standard as sumption in economic theory. Even though many of the real world markets being modeled (e.g., insurance markets, financial markets, and labor markets) were recognized as having im portant information imperfections, it was not subject to debate until fairly recently. Stigler's (1961) article on "The economics of informa tion" is recognized as a pathbreaking article on the formal modeling of imperfect information.

Incomplete information is a type of imperfect information. For example, one or more parties to a transaction may lack knowledge of some aspect of the transaction that affects their payoffs. When consumers have incomplete information, markets which would be perfectly competitive otherwise have the features of MONOPOLISTIC COMPETITION and firms have some degree of MARKET POWER. ASYMMETRIC INFORMA TION is another type of imperfect information. For example, one party to a transaction may have valuable knowledge of some aspect of the trans action that another party does not. Information asymmetries are an important source of infor mation imperfections which may in turn create problems of MORAL HAZARD and ADVERSE SELECTION.

An early article on the effects of imperfect information on the standard conclusions of eco nomic theory is Rothschild and Stiglitz (1976). They examined the effects of imperfect infor mation on competitive insurance markets and found that under plausible conditions an EQUI LIBRIUM did not exist in those markets in the presence of imperfect information. Although their focus was on insurance markets, their results also had implications for financial markets and labor markets. In the context of credit markets, Stiglitz and Weiss (1981) show how equilibrium in credit markets can be char acterized by rationing in the presence of imper fect information.

One of the implications of Rothschild and Stiglitz (1976) was recognizing imperfect infor mation as a source of market inefficiency. An other was recognizing that specific institutions in markets may be responses to the difficulties of handling problems of imperfect information. This second implication has had a profound influence on recent developments in what has come to be called the "new institutional econom ics." The approach taken by the imperfect infor mation strand of the new institutional economics is associated primarily with the work of Akerlof and Stiglitz. It is rigorous, specifying assump tions and equilibrium solutions, and making sharp distinctions between different types of information problems. Stiglitz (1986) utilizes the imperfect information theory of institutions to provide an explanation of the existence of certain agrarian institutions as substitutes for incomplete insurance markets, credit markets, and futures markets.

Screening and SIGNALING are two methods individuals use in attempting to overcome prob lems of imperfect information that arise due to adverse selection. Stiglitz (1975) develops a theory of screening which he uses to analyze the allocation of resources to education. Screening serves to sort individuals on the basis of their qualities. Stiglitz finds that screening in education has productivity returns but it increases inequality, creating a trade off between distributional and efficiency consider ations. Leland and Pyle (1977) develop a signal ing model where entrepreneurs seek financing for projects whose true qualities are known only to them. An entrepreneur's willingness to invest in his or her own project is seen as a signal of the project's quality. They conclude that the value of the firm increases with the share of the firm owned by the entrepreneur, and suggest that financial intermediation is a natural response to asymmetric information.

See also *adverse selection; asymmetric information; moral hazard*

Bibliography

Leland, H. and Pyle, D. (1977). Informational asymmet ries, financial structure and financial intermediation. *Journal of Finance*, **32**, 371 87.

Rothschild, M. and Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the eco nomics of imperfect information. *Quarterly Journal of Economics*, **90**, 629 50.

Stigler, G. J. (1961). The economics of information. *Jour nal of Political Economy*, **69**, 213 25.

Stiglitz, J. (1975). The theory of "screening," education and the distribution of income. *American Economic Review*, **65**, 283 300.

Stiglitz, J. (1986). The new development economics. *World Development*, **14**, 257 65.

Stiglitz, J. and Weiss, A. (1981). Credit rationing in markets with imperfect information. *American Economic Review*, **71**, 393 410.

income effect

Robert E. McAuliffe

The income effect refers to the change in a consumer's real income (or purchasing power) when the price of a product changes. When the price of a product increases, those consumers who purchase the product have lost purchasing power and will tend to buy less of all goods – including the good whose price rose. The strength of this income effect depends on the importance of the product in the consumer's budget and the availability of SUBSTITUTES. Thus an increase in the price of a specific manu facturer's product of, say, chicken will have less effect on consumers than an increase in the price of all chicken, and this will have less effect than an increase in the price of all food products.

Just as consumers are worse off from a price increase, a price decrease will make consumers better off by increasing their purchasing power. For most products, the income effect is positive – that is, consumers tend to purchase more of the good as their incomes rise. This holds for NORMAL GOODS and for luxury goods, but the income effect is negative for INFERIOR GOODS. The income effect normally reinforces the SUBSTITUTION EFFECT, causing demand functions to slope downward (*see* DEMAND FUNCTION).

See also *law of demand*

Bibliography

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Publishing.

Varian, H. R. (1992). *Microeconomic Analysis*, 3rd edn. New York: Norton.

income elasticity

Gilbert Becker

The income elasticity is defined as the percent age change in quantity demanded given a small percentage change in income, and is calculated by the ratio

$$\varepsilon_I = \frac{\Delta Q}{Q} \Big/ \frac{\Delta I}{I} = \frac{\Delta Q}{\Delta I} \times \frac{I}{Q}$$

where the symbol Δ indicates a change in quan tity (Q) or income (I). Its measure may take on a positive or negative value and is of considerable importance to managers. For most goods an increase in income raises demand and thus ε_I has a value greater than zero. For some of these goods, formally classified as luxury goods, the change in demand is greater than the income change and ε_I takes on a value greater than 1. For a few goods, such as bus transportation, increased income causes substitution to alterna tive goods and the value of ε_I is negative. These goods are identified as INFERIOR GOODS. The degree of consumer sensitivity to income changes is important in anticipating the periodic growth and decline in demand for a good. Products with a large positive value for income elasticity will find demand growth exceeding the growth rate of income in the economy. These products are also more susceptible to sharp declines in demand during recessions. Suppliers of inferior goods, on the other hand, should expect a decline in demand for their product in the long run as consumer income grows.

See also *elasticity*

Bibliography

Bordley, R. and McDonald, J. (1993). Estimating aggre gate automotive income elasticities from the population

income-share elasticity. *Journal of Business and Eco
nomic Statistics*, **11**, 209 14.

Maurice, C. and Thomas, C. (2002). *Managerial Econom
ics*, 7th edn. Boston: McGraw-Hill Irwin.

Nicholson, W. (2000). *Intermediate Microeconomics and its
Application*, 8th edn. New York: Harcourt.

incomplete contracts

Lidija Polutnik

Incomplete contracts are a response to bounded rationality which results from the inability of people to exactly foresee and articulate all pos sible contingencies or because the cost mini mization requires the contract to be incomplete (*see* TRANSACTIONS COSTS). Simon (1951) analyzed the consequences of bounded rational ity for the employment relationship. A complex long term relationship between the employee and the employer is most commonly defined as an incomplete contract and is a rational response to human limits of bounded rationality and un certainty. When contracts are incomplete, a pos sibility for opportunistic behavior exists because incentives are not properly aligned, and this leads to problems of imperfect commitment.

Imperfect commitment can lead to an incen tive for one of the parties of the contract to prefer renegotiation of the contract *ex post*. For example, when the contract fails to specify how the price should change over time as a result of a change in costs, the buyer may refuse to take the supplier's word for the change in costs, and a risk averse buyer will find a contract that does not specify prices with certainty unattractive (*see* RISK AVERSION). Fear of reneging may make contracting parties engage in costly behavior aimed at protecting themselves to assure com pletion of a contract or decide not to enter into a contract in the first place.

As a result of contractual incompleteness, a hold up problem may occur. Certain transac tions may necessitate that specific investments be made. When this investment is made, the value of the input in its present use is much higher than in its next best use, which creates a basis for opportunistic behavior. One of the parties may refuse to abide by the contractual terms after the specific investment has been made. The party whose contract terms have worsened as a result of contractual incomplete ness is said to be held up. This inability to commit parties to postcontractual behavior may result in fewer asset specific investments being made (*see* ASSET SPECIFICITY).

Williamson (1979) showed that the problem of commitment and opportunistic behavior is lessened when firms engage in long run relation ships with frequent interaction and the degree of asset specificity of the investment made by the parties is low. There is a variety of contractual means for reaching a commitment when parties engage in complex transactions that last over a period of time and contracts are incomplete. A hold up problem may be avoided and oppor tunism mitigated if a single firm owns highly specialized assets (*see* VERTICAL INTEGRA TION). Relational contracts are a common way of trying to resolve unspecified contingencies should they occur. They are abstract in nature but do necessitate specification of an authority or governance structure. For example, a supervisor has discretionary power over an employee's as signment of tasks. Also important in practice are "implicit contracts." These are defined by a set of shared expectations about the behavior or a relationship. They are not enforceable in a court of law but are designed to be self enforcing. Rosen (1985) provides a useful survey of studies pertaining to the implicit nature of employment contracts and explanations for different phe nomena, such as wage rigidity, human resource policies in a firm and their objectives for employ ment security, recruiting, etc.

A noncontractual way of achieving commit ments and aligning incentives when contracts are incomplete is with reputations. Reputational effects can sustain contracts until the near end and save the firm the costs of writing a complete contract. In the employment relationship the firm reneging on a promise of a wage increase damages its reputation and reduces the future possibility of an equally profitable relationship with its employees. Kreps (1990) provides numerous examples of corporate culture being employed to enforce expected behaviors in a corporation.

See also *asset specificity; Coase theorem; game theory; transactions costs*

Bibliography

Brickley, J. A., Smith, C. W., and Zimmerman, J. L. (2000). *Managerial Economics and Organizational Architecture*. Chicago: McGraw-Hill.

Carlton, D. W. and Perloff, J. M. (2000). *Modern Industrial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Kreps, D. (1990). Corporate culture and economic theory. In J. Alt and K. Shepsle (eds.), *Perspectives on Positive Political Economy*. Cambridge, MA: Cambridge University Press, pp. 90 143.

Milgrom, P. and Roberts, J. (1992). *Economics, Organization and Management*. Englewood Cliffs, NJ: Prentice-Hall.

Rosen, S. (1985). Implicit contracts: A survey. *Journal of Economic Literature*, **23**, 1144 75.

Simon, H. A. (1951). A formal theory of the employment relationship. *Econometrica*, **19**, 293 305.

Williamson, O. (1979). Transaction-cost economics: The governance of contractual relations. *Journal of Law and Economics*, **22**, 233 61.

Williamson, O. (1985). *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. New York: Free Press.

increasing returns

Robert E. McAuliffe

There are increasing returns to an input in production when, as incremental units of that input are added, the resulting increases in output are rising. Increasing returns typically occur in the early stages of production as the first units of a variable input are added, but, as more units are added, DIMINISHING RETURNS set in (*see* LAW OF VARIABLE PROPORTIONS). For example, the increases in output from adding the first workers to a plant should rise as those workers assist each other in production, so the MARGINAL PRODUCT of labor would be rising. Eventually, additional workers will not be as productive and their marginal product will be falling. As with diminishing returns, increasing returns refer to the short run when other factors of production are held constant. When all factors of production are changed, the issue is one of ECONOMIES OF SCALE (or of returns to scale).

Bibliography

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Publishing.

indifference curves

Gilbert Becker

Indifference curves are used to describe the preferences of a consumer who is faced with an array of market baskets, or sets of goods, from which to choose. One indifference curve represents all of the different combinations of goods which give the consumer equal TOTAL UTILITY. Typically, market baskets containing different quantities of two goods are examined. Ordinarily, as the amount of either of the goods is reduced, utility is lost. As a result, the quantity of the second good must be increased in order to compensate for the loss and to leave the consumer indifferent between the initial and new positions. For example, a consumer may indicate that a combination of four pairs of sneakers and one warm up suit per year gives the same overall satisfaction as do three pairs of sneakers and two warm up suits or one pair of sneakers and four warm up outfits. These three combinations would be part of the same indifference curve for that consumer.

As it is generally assumed that more is preferred to less, any combination having more of at least one of the goods and no less of the second good would generate more total utility. Thus, for example, four pairs of sneakers in combination with three outfits would yield more total utility than the first combination identified above (and, by transitivity, more than the other two combinations as well). When other combinations having the same total utility as this new point are identified, a second indifference curve is created. The complete set of such curves is known as an **indifference map**.

It is important to recognize that the indifference curve represents the consumer's evaluation of these combinations without regard to income or the prices of the two goods. As such, the indifference map identifies a complete ranking of these combinations of goods solely on the basis of the consumer's tastes. The various combinations of goods which the consumer can afford and that set which leads to the highest total utility possible require the analysis of the consumer's BUDGET CONSTRAINT as well (see UTILITY MAXIMIZATION).

A diagram of an indifference curve is drawn with the y axis measuring the quantities of one good (Y) and the x axis measuring the quantities of the second good (X). The different combinations that are equivalent in the eyes of the consumer are combined to form a curve, which ordinarily is negatively sloped. The slope of the indifference curve indicates the amount of good Y which the consumer is willing to trade in order to receive one extra unit of good X. This is identified as the MARGINAL RATE OF SUBSTITUTION of X for Y. Each indifference curve is typically bowed inward toward the origin due to the law of DIMINISHING MARGINAL UTILITY, which causes the marginal rate of substitution between the two goods to decline. That is, as the consumer sacrifices units of Y to consume more of good X, he is less willing to do so the more X is consumed. Indifference curves that are farther from the origin represent higher levels of overall utility, and are therefore more desirable.

Indifference curve analysis can be used to demonstrate how an ADVERTISING campaign for a product works to alter consumers' marginal rates of substitution and thus favorably alters their choices toward greater purchases of the good being promoted. Alternatively, the analysis can be used to show how a pricing strategy which lowers the price of one good causes greater consumption of that good through the choice of a new combination on the consumer's indifference map. Moreover, both the INCOME EFFECT and the SUBSTITUTION EFFECT of the price change can be demonstrated.

In addition, indifference curves can be used to examine investment decisions by managers. Combinations of risk and rates of return (in the place of goods X and Y) can be used to create indifference curves which describe varying degrees of RISK AVERSION. Since risk is an economic "bad" (consumers prefer less risk), the indifference curves slope *upward* when expected rates of return are plotted on the vertical axis and against risk on the horizontal axis. This means that if consumers are to increase the riskiness of their portfolio, they would need to be compensated by a higher expected rate of return to remain indifferent. One application is portfolio analysis (see CAPITAL ASSET PRICING MODEL).

Finally, important advances have been made by Lancaster (1971), who used indifference curves to describe consumer preferences in terms of product characteristics. By identifying consumers' tastes toward combinations of features that a product may hold, *target markets* can be established, the development and introduction of new products can be implemented, and marketing strategies concerning the BUNDLING of options and other forms of PRODUCT DIFFERENTIATION can be developed (see PRODUCT ATTRIBUTES MODEL).

Bibliography

Lancaster, K. (1971). *Consumer Demand: A New Approach*. New York: Columbia University Press.

Nicholson, W. (2000). *Intermediate Microeconomics and its Application*, 8th edn. New York: Harcourt.

industry life cycle

Robert E. McAuliffe

This is the extension of the PRODUCT LIFE CYCLE concept to the industry. When a new product (or product category) has been created, the industry is in the **introduction** phase of the cycle where the market is small but growing and risk is high because the market is not yet fully established. Firms tend to have high ADVERTISING expenditures in this stage as they inform consumers about the new product. In addition, firms may be more vertically integrated if markets for supporting services and parts have not developed (see VERTICAL INTEGRATION). For example, in the early years of the computer industry there were few independent producers of peripherals such as disk drives or software, so companies such as IBM and later Digital had to

produce these products themselves. During the **growth** phase of the cycle, the market grows rapidly as new firms enter the industry (*see* ENTRY). New product development, innovations, and marketing methods occur during this period as firms compete for higher profits and market share. Eventually as the market becomes saturated, growth slows and the industry reaches the **maturity** phase of the industry cycle where profits stabilize and firms tend to compete more on price rather than through PRODUCT DIFFERENTIATION. Following this phase, the industry enters the stage of **decline** with a contracting market and higher EXIT rates among firms. The industry structure tends to consolidate as larger firms acquire weaker rivals, COMPETITION focuses on price, and costs and industry profits are falling.

As Porter (1980) observed, there are several criticisms of the life cycle approach. Every industry or product will not necessarily pass through each stage of the cycle, nor is it always clear where a specific industry may be in the life cycle. The evolution of the industry and competition among firms may vary between industries rather than follow the course predicted by the life cycle. However, the industry life cycle does offer interesting dynamic predictions regarding entry, exit, competition, and profits over time (Greer, 1992; *see* DECLINING INDUSTRY).

Bibliography

Greer, D. F. (1992). *Industrial Organization and Public Policy*, 3rd edn. New York: Macmillan.

Jacquemin, A. (1987). *The New Industrial Organization: Market Forces and Strategic Behavior*. Cambridge, MA: MIT Press.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.

inelastic demand

Gilbert Becker

Inelastic demand is the situation where consumer reaction to a change in price is small. Specifically, it is defined as the case where the percentage change in quantity demanded is less than the percentage change in price, or $\varepsilon_p < 1$, where ε_p is the price elasticity of demand (*see* ELASTICITY). When a class of products has few or no SUBSTITUTES, such as gasoline, market demand tends to be inelastic. As the product is more narrowly defined though, and a specific brand of gasoline is considered, more substitutes are available and its demand becomes more elastic. In addition, goods considered to be necessities, such as food or prescription medications, and goods that play a small role in consumers' budgets, such as salt, also tend to have demands that are inelastic. In all cases of inelastic demand, price increases will increase TOTAL REVENUE and price decreases will lower revenues. Moreover, in cases where consumers of the same product (e.g., movie tickets) can be separated into different groups (e.g., children, adults, senior citizens) and charged different prices, elasticity plays a critical role in the pricing strategy. In order to maximize profits, managers should charge the highest price to the group with the most inelastic demand, and lower prices to groups whose demand is more elastic (*see* PRICE DISCRIMINATION; PROFIT MAXIMIZATION).

Bibliography

Maurice, S. and Thomas, C. (2002). *Managerial Economics*, 7th edn. Boston: McGraw-Hill Irwin.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

inferior goods

Nikolaos Vettas

A good is called inferior if consumers choose to buy less of this good when their REAL INCOME is higher. An example that is often used to illustrate this behavior is potatoes: as people's income rises, they may choose to consume fewer potatoes (and more of some other food). Goods that are not inferior are called NORMAL GOODS. Of course, a good may be normal at some income levels and inferior at other levels of income.

For a DEMAND FUNCTION to have positive slope, it is necessary (but not sufficient) that the good is inferior. The demand curve for an inferior good may be either increasing in price ("Giffen good") or decreasing in price, depending on whether the INCOME EFFECT or the SUBSTITUTION EFFECT dominates (*see* DEMAND CURVES).

Bibliography

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Publishing.

inflation

Robert E. McAuliffe

Inflation measures the rate of change in the general level of prices in the economy. The inflation rate is calculated as the percentage change in a price index (such as the CONSUMER PRICE INDEX (CPI), the producer price index, or the GDP deflator) over a given period of time. For example, if the CPI at the end of May 2002 is 100 and its value for the end of June 2002 is 101, then the inflation rate for the month would be

$$\text{Inflation} = \left[\frac{(\text{CPI}_{\text{June}} - \text{CPI}_{\text{May}})}{\text{CPI}_{\text{May}}}\right] \times 100\%$$
$$= \left[\frac{(101 - 100)}{100}\right] \times 100\% = 1\%$$

Since a company's revenues and costs may adjust differently when inflation occurs, managers should consider the effects inflation may have on profits and may need to estimate its effects on costs and revenues separately for accurate forecasts. A firm might use the producer price index to measure inflation changes in its costs and the CPI to measure inflation in revenues. A subset of either price index may be preferred if specific elements of price change in the company's costs or revenues were more suitable. Thus a clothing retailer might be more interested in the price changes in clothing specifically than in changes in the overall index of consumer or producer prices and would choose that specific category of items in the price index.

Bibliography

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Publishing.

internal organization of the firm

Lidija Polutnik

Coordination of tasks and a division of work can be organized quite differently among different firms. One of the first methods of organizing activities within a firm was the **functional organization**. Firms were organized in a centralized fashion according to the functions they performed; for example, separate departments for finance, marketing, production, purchasing, etc. were formed. This type of internal organization of the firm permitted specialization of labor, and the guiding principle for the formation of departments was technical efficiency. However, functional organizations were not very capable of accommodating the needs of large, multiproduct corporations that performed their activities in several geographic regions.

The expansion of product lines in the 1920s and increasing multiproduct production by firms led to a new organizational form of a firm: the **multidivisional** (or **M form**) organization. Du Pont was the first company to introduce this new form as a result of difficulties in producing goods as diverse as paints, fertilizers, and vegetable oils. New divisions were created within companies, each of which was responsible for a product, geographical location, market, or technology. These divisions may then be organized functionally. Multidivisional organization permits less specialization, but it makes each

division and its management responsible for PROFIT MAXIMIZATION. Multidivisional or ganization evolved as a result of minimizing not only production costs but TRANSACTIONS COSTS as well (*see* PRINCIPAL–AGENT PROB LEM). For example, this type of organization allows for better coordination of activities: div ision managers are more knowledgeable and better informed about their particular product lines and their incentives are better aligned with the overall performance objectives of the com pany than under functional organizations.

Bibliography

Brickley, J. A., Smith, C. W., and Zimmerman, J. L. (2000). *Managerial Economics and Organizational Archi tecture*. Chicago: McGraw-Hill.
Carlton, D. W. and Perloff, J. M. (2000). *Modern Indus trial Organization*, 3rd edn. Reading, MA: Addison-Wesley.
Milgrom, P. and Roberts, J. (1992). *Economics, Organiza tion and Management*. Englewood Cliffs, NJ: Prentice-Hall.

internal rate of return

Robert E. McAuliffe

The internal rate of return (IRR) is the discount rate which makes the NET PRESENT VALUE of an asset equal to its current cost. Formally, if an asset which costs P_0 today will generate cash flows of C_i in each period i for T periods in the future, then the IRR is that value of r^* which solves

$$\sum_{i=1}^{T}\left[\frac{C_i}{(1-r^*)^i}\right] - P_0 = 0$$

The IRR essentially measures the expected rate of return from an investment, and this can be compared with some hurdle rate for CAPITAL BUDGETING decisions. Based on this criterion, an investment should be made if its IRR exceeds the cost of capital. The IRR may not be unique if future cash flows alternate between positive and negative values, and the ranking of alternative investments by this criterion may not corres pond with rankings from the net present value method (see Edwards, Kay, and Mayer, 1987;

Brigham and Gapenski, 1997). Fisher and McGowan (1983) assert that the IRR is the ap propriate measure of the economic rate of return, but Edwards et al. suggest that adjusted accounting data could be used in economic an alysis.

See also *cash flow*

Bibliography

Brigham, E. F. and Gapenski, L. C. (1997). *Intermediate Financial Management*, 5th edn. New York: Inter national Thomson Publishing.
Edwards, J., Kay, J., and Mayer, C. (1987). *The Economic Analysis of Accounting Profitability*. Oxford: Oxford University Press.
Fisher, F. M. and McGowan, J. J. (1983). On the misuse of accounting rates of return to infer monopoly profits. *American Economic Review*, **73**, 82 97.
Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.
Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Pub lishing.

isoquant-isocost curves

Gilbert Becker

These curves form the basis for the analysis of the choice of the optimal combination of inputs to be used in the production of any good. An isoquant curve identifies the set of all combin ations of inputs which yield the same level of total output. As such, it is the production analy sis counterpart to consumption INDIFFERENCE CURVES. Since less of any one input typically decreases total output, more of the second input is needed as compensation in order to maintain constant total output (*see* MARGINAL RATE OF TECHNICAL SUBSTITUTION). Because of this, isoquant curves are ordinarily negatively sloped, although in circumstances where too much of one input is used the result may be a negative marginal productivity from that input (*see* DIMINISHING RETURNS). At this point the isoquant curve's slope becomes positive.

One example comes from agricultural produc tion where it has been demonstrated by Heady

(1957) that various combinations of different per acre amounts of two fertilizers led to an equal per acre yield of a given crop. For any one isoquant, greater amounts of one input (phosphate) will require lesser amounts of the second (nitrogen). Increases in both inputs will, up to some level, cause greater total output. A series of isoquant curves, each identifying different levels of output, can be identified and an isoquant map can be constructed. Similar studies have used engineering data to examine the production pro cess for automobiles, solar heating, and other industries.

An isocost line identifies the financial con straint which the firm faces in terms of purchas ing inputs to be used in the production process. It includes both the prices (P) of the inputs in question and the total dollar amount of funds (F) available for inputs. This is described in the following equation for the case of two inputs:

$$F = P_a \times Q_a + P_b \times Q_b$$

where Q designates the quantity of each input (a,b) used (*see* BUDGET CONSTRAINT for a description of the equivalent limitation facing the consumer).

The isocost line identifies the different pos sible maximum combinations of inputs which the firm is able to buy with a fixed amount of available funds. Its slope, measured by the ratio of the input prices, represents the trade off which the market requires of the firm. That is, it indicates the amount of one input which must be sacrificed in order to purchase an extra unit of the second input.

The tangency point between the firm's isocost line and the highest achievable isoquant curve indicates the optimal combination of the two inputs to be selected. This combination maxi mizes the total output which can be produced from a given expenditure level on inputs. Equivalently, the combination minimizes the total cost of producing a given level of output. Since a firm's profits are equal to its TOTAL REVENUE minus its TOTAL COSTS, cost mini mization through the selection of an ideal com bination of inputs is essential for a firm striving to achieve PROFIT MAXIMIZATION.

See also *utility maximization*

Bibliography

Browning, E. and Zupan, M. (2004). *Microeconomics: Theory and Applications*, 8th edn. New York: John Wiley.
Heady, E. (1957). An econometric investigation of the technology of agricultural production functions. *Econometrica*.
Maurice, S. and Thomas, C. (2002). *Managerial Econom ics*, 7th edn. Boston: McGraw-Hill Irwin.
Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

J

junk bonds

Vickie L. Bajtelsmit

Junk bonds are high risk debt securities that are rated Ba (or lower) by Moody's Investment Services, Inc. and BB (or lower) by Standard and Poor's Corporation. Standard and Poor's Bond Guide defines this rating as ''predominantly speculative with respect to capacity to pay interest and repay principal in accordance with the terms of the obligation.'' Because many institutional investors have restrictions against investing in speculative securities, low rated bonds are termed ''below investment grade.'' To offset the greater risk exposure to investors, low grade bonds carry yields that are at least 3 percent greater than that of high quality corporate debt.

Prior to the 1970s, publicly traded bonds were not issued with below investment grade ratings, although they could be subsequently downgraded. During the 1980s, Drexel Burnham Lambert (DBL), a large investment banking firm, increased interest in this type of investment by pioneering the use of newly issued junk bonds as a source of financing for takeovers and mergers. DBL convinced investors that junk bonds offered a high return without commensurate risk and also offered to provide liquidity to the market by promising to buy back bonds from their customers on demand. However, the economic environment of the late 1980s and early 1990s resulted in many defaults and, ultimately, the failure of DBL. The result has been reduced investor interest in this type of security.

Bibliography

Gruber, M. J. (1991). *Modern Portfolio Theory and Investment Analysis*, 4th edn. New York: John Wiley.
Kolb, R. W. (1992). *Investments*, 3rd edn. Miami: Kolb.

K

kinked demand curves

Gilbert Becker

These demand functions demonstrate the rela
tion between the price of a firm's product and its
quantity demanded when taking into account
specific reactions by rival firms. The LAW OF
DEMAND indicates that, all else constant, as the
price of one firm's product is reduced, its quan
tity demanded will rise. This generates a
demand curve which typically is assumed to be
continuous. Under OLIGOPOLY, rival firms'
reactions can affect the perceived shape of the
demand curve. In MONOPOLY, PERFECT
COMPETITION, and MONOPOLISTIC COM
PETITION, rivals' reactions do not matter.

Numerous models have been developed
which apply different assumptions concerning
rivals' reactions to a price change by one firm.
One of the most widely known models is that
developed by Sweezy (1939), in which rivals are
expected to react to a firm's price increase by not
changing their own price but are expected to
match any price decrease made by the first
firm. The strategy by the rivals here is designed
to increase their market share (for the case of
price increases by the first firm) and protect their
market share (for the case of a price decrease). If
the first firm expects such asymmetry in the
reactions by its rivals, it would anticipate a
sharp drop in unit sales for any price increase
that it initiated. That is, it expects the demand
for its product to be *elastic* for a price rise. At the
same time, a price cut by this firm would be
expected to bring a small increase in unit sales,
that is, a more INELASTIC DEMAND for a de
crease in price. Hence, the complete demand
curve for such a firm contains a distinct kink at
its current price level.

Kinked demand curves were first proposed as
models of oligopoly. They attempted to explain a
perceived inflexibility in oligopoly pricing rela
tive to that which occurred in more competitive
markets. As the firm that considers initiating a
price change anticipates a harmful response from
its rivals, these expectations may deter it from in
fact changing price. The theory has maintained
its popularity in part through its ability to ex
plain, at times, the pricing behavior in some
markets, especially those for homogeneous
goods where price becomes a very important
strategic variable. Evidence by Stigler (1947),
Primeaux and Bomball (1974), and others rejects
the existence of a kink in the demand curves of
firms in several oligopoly markets. Another con
sidered failure is the inability of the model to
explain the current price of the firm. These and
other shortcomings indicate that kinked demand
curve models do not form a general theory for
oligopoly.

See also *demand function; demand curves; game
theory*

Bibliography

Primeaux, W. and Bomball, M. (1974). A reexamination
of the kinky oligopoly demand curve. *Journal of Polit
ical Economy*, **82**, 851 62.

Schiller, B. (2003). *The Micro Economy Today*, 9th edn.
New York: Irwin McGraw-Hill.

Stigler, G. (1947). The kinky oligopoly demand curve and
rigid prices. *Journal of Political Economy*, **55**, 432 49.

Sweezy, P. (1939). Demand under conditions of oligop-
oly. *Journal of Political Economy*, **47**, 568 73.

L

law of demand

Gilbert Becker

A fundamental principle of economics states that an inverse relationship exists between a change in the price of any product and the resulting change in its quantity demanded in the market. More formally the law of demand states that, CETERIS PARIBUS, as the price of a good rises, its quantity demanded falls, and as the price of a good falls, its quantity demanded rises in the market. It is important to note that the motivating force behind the change in consumer behavior here is a change of price rather than, and without, any other change occurring in variables such as consumer tastes and income, which also make up the DEMAND FUNCTION.

The price change influences the adjustment in quantity demanded through two separate effects. The INCOME EFFECT is that portion of the change in quantity demanded which is due to the change in consumer purchasing power resulting from the price change. For example, when the price of a product decreases, the existing (and unchanged) level of consumer income now has greater real purchasing power. This makes consumers of the good in question richer and induces an increase in the number of units purchased. The SUBSTITUTION EFFECT recognizes that in the calculus of consumer decision making, individuals are also concerned with the price of a product relative to the prices of any SUBSTITUTES that are available. Since the ceteris paribus condition in the law holds these other prices constant, a fall in the absolute price of a good causes a decrease in its price relative to all others. This makes the good in question more attractive to consumers, who then substitute away from other goods toward this good. Both effects normally reinforce the inverse relationship between price and quantity demanded.

Moreover, the law is consistent with, and in part derived from, the concept of DIMINISHING MARGINAL UTILITY, which indicates that successive units purchased of any good, in a given time period, offer less value to a consumer. As such, consumers are less likely to be willing to pay as much for additional units that offer lower utility, and the per unit price must be reduced to coax additional sales. An example of this can be seen at many fast food restaurants where the price per unit of the medium size soft drink is considerably lower than the smaller size, as the MARGINAL UTILITY of the extra soda diminishes. In fact, many of these businesses offer free refills (i.e., price equals zero) for initial purchases of the largest size.

The near universal applicability of this consumer behavior to a wide array of goods has elevated the principle to the status of an economic law. Even the demand for goods that are absolute necessities, such as water, tends to follow this law. Studies have shown that when people are faced with higher water prices, while they do not necessarily drink less water they will economize on its other uses. Also, what often appears to be a counter example to the law is actually a situation in which one or more of the other variables in the demand function have changed and the ceteris paribus condition has been broken.

For some products, the existence of NETWORK EXTERNALITIES may also influence the relation between the product's price and quantity demanded (see Shapiro and Varian, 1999). Here an individual consumer's utility from purchasing a good depends on the number of other consumers who may buy that good. For example, as cellular phones become popular, a

lower price may induce more sales due to a "bandwagon effect" as well as the income and substitution effects. For a given individual, the lower price for phones may create the expectation that other consumers will purchase phones as well. This increases the phones' utility and hence the quantity demanded at the lower price.

Economic laws are not always obeyed and thus the law of demand is not without exceptions. One such example is when a negative network externality or "snob effect" occurs. "Snobs," for whom relatively exclusive ownership of the product is desirable, may increase their purchases of the good as its price rises if the higher price makes the good more exclusive. Here the individual's demand shows a positive relation between the product's price and the quantity demanded. The market demand for this product sometimes, but not always, continues to maintain the inverse relation indicated by the law.

A second example where the law of demand may be violated concerns markets where new goods are involved. Bagwell and Riordan (1991) indicate that a higher price for a new product may induce increased sales as it sometimes acts as a signal of higher quality in markets where information is imperfect (*see* ASYMMETRIC INFORMATION).

Managers should be aware that the law indicates that a price change will ordinarily have an important impact on a firm's sales volume and TOTAL REVENUE, and thus its profitability. Although the law identifies the *direction* of the change in quantity demanded resulting from a price change, it does not indicate the size of the change. This size change is central to the overall impact of the price change on revenues (*see* ELASTICITY; INELASTIC DEMAND).

Bibliography

Bagwell, K. and Riordan, M. K. (1991). High and declining prices signal product quality. *American Economic Review*, **81**, 224 39.

Browning, E. and Zupan, M. (2004). *Microeconomics: Theory and Applications*, 8th edn. New York: John Wiley.

McConnell, C. and Brue, S. (2002). *Microeconomics*, 15th edn. Boston: McGraw-Hill Irwin.

Shapiro, C. and Varian, H. (1999). *Information Rules*. Boston: Harvard Business School Press.

law of variable proportions

Robert E. McAuliffe

Sometimes referred to as the law of DIMINISHING RETURNS, the law of variable proportions is concerned with the effect of changes in the proportion of the factors of production used to produce output. As the proportion of one input increases relative to all other inputs, at some point there will be decreasing marginal returns from that input. Adding more units of an input, holding all other inputs constant, will at some point cause the resulting increases in production to decrease, or equivalently, the MARGINAL PRODUCT of that input will decline. Among the inputs held constant is the level of technology used to produce that output. This is an empirical law and is therefore a generalization about the nature of the production process and cannot be proven theoretically (see Friedman, 1976; Stigler, 1966). Applied to management, Friedman argues that the law of variable proportions requires firms to produce by using inputs in such proportions that there are diminishing average returns to each input in production.

Bibliography

Friedman, M. (1976). *Price Theory*, 2nd edn. Chicago: Aldine.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Stigler, G. J. (1966). *The Theory of Price*, 3rd edn. New York: Macmillan.

learning curve

Robert E. McAuliffe

The learning curve refers to the reduction in AVERAGE TOTAL COST which occurs as workers gain experience from producing a product over time, and for this reason it is also called the experience curve. Unlike ECONOMIES OF

SCALE where long run average costs decrease when more output is produced per period of time, the learning curve shows the reduction in average costs arising from the total accumulated volume of production to date. Therefore if a firm produced 2 million units per month at MINIMUM EFFICIENT SCALE and the average cost per unit was $10, if there were learning effects in production, they would cause the average cost curve to shift down over time. In this example, the firm might find that its average costs of production fell to $8 per unit once the firm had produced 24 million units over a year, even though production remained at 2 million units each month.

One of the first theoretical treatments of learning effects was provided by Arrow (1962), and learning curves have been estimated for a variety of industries (see Ghemawat, 1985, for a survey). When significant learning effects exist, they can confer strategic advantages (*see* FIRST MOVER ADVANTAGES) to those established firms which increase their production volumes quickly to reduce their costs (Porter, 1980). To take advantage of the learning curve, managers should set prices for new products below the level that would only maximize current period profits, recognizing that future costs will be lower and profits higher.

This aspect of pricing with a learning curve makes it difficult to establish whether a firm has engaged in PREDATORY PRICING, because a firm may price below average total cost anticipating lower costs in the future from learning effects. Since this form of pricing is a legitimate business practice and leads to more efficient production, it is difficult to determine whether pricing below average cost is actually predatory in nature (see Carlton and Perloff, 2000).

Some care is required when estimating learning curves because the effects occur over time, but as time passes other factors such as FACTOR PRICES which affect average costs will also change. A simple but common representation of the learning curve would be:

$$AC_t = AC_0 \times CV_t^{\lambda} \times e^{u_t} \qquad (1)$$

where AC_t is the *real* average cost per unit of production in period t, AC_0 is the real average cost in the initial period of production, CV_t

represents the cumulative volume of output produced up to period t, λ is the ELASTICITY of average costs with respect to volume, and u_t is an error term (with e the natural exponent). Since equation (1) relates current real average costs of production to the initial cost and the total volume produced, it omits the costs of factors of production which, if they have changed during the sample period, will cause estimates of the learning curve to be biased (see Berndt, 1991). In fact, Berndt shows that unless the effects of changes in input prices can be captured by an appropriate deflator and there are constant returns to scale, estimates of the learning curve based on (1) will be biased.

If we ignore the biases mentioned above, how would a manager estimate the learning curve? Taking logarithms of both sides of (1) yields:

$$\ln(AC_t) = \ln(AC_0) + \lambda \times \ln(CV_t) + u_t \quad (2)$$

which can be estimated by LINEAR REGRESSION. As Berndt observed, it is important to use general price deflators to obtain real unit average costs, since a price deflator for the industry will already include learning effects and will therefore mask the cost reductions we wish to estimate. Once λ has been estimated, average real costs will decrease according to

$$AC_{\text{new}} = (2^{\lambda}) \times AC_{\text{old}} \qquad (3)$$

when total volume doubles. For example, if the estimated value of λ were -0.25, then costs would decrease by 25 percent of their previous level when production volume doubled and the learning curve would have a 75 percent slope.

Bibliography

Arrow, K. J. (1962). The economic implications of learning by doing. *Review of Economic Studies*, **29**, 153 73.

Berndt, E. R. (1991). *The Practice of Econometrics*. Reading, MA: Addison-Wesley.

Carlton, D. W. and Perloff, J. M. (2000). *Modern Industrial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Ghemawat, P. (1985). Building strategy on the experience curve. *Harvard Business Review*, **63**, 143 9.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.

lemons market

Vickie L. Bajtelsmit

In information theory, a "lemons market" is a market in which the degree of ASYMMETRIC INFORMATION between buyers and sellers is very high, and, in the extreme, may result in market failure. Akerlof (1970) provided an intuitive and logical explanation of this theoretical result by reference to a used car market. In a used car market, there may be cars that are of good quality and those that are "lemons." Since sellers are generally better informed than buyers regarding the quality of the car and buyers are not easily able to discern car quality, buyers will be unwilling to pay the "good car" price for a car of uncertain quality, so they value a car at the average price. In the extreme, owners of good cars will be unwilling to sell their cars at the prevailing price and, ultimately, the only cars in the market will be those that no one wants and the market will fail.

The solution to this problem requires that market participants act to reduce the level of information asymmetry. For example, buyers might invest in additional information by hiring experts or gaining expertise themselves. Sellers may attempt to better convey information, although they will have to bear some costs to make their claims of value believable to potential buyers (*see* SIGNALING).

The logic and intuition used in the used car example has been extended to several other markets that exhibit imperfections due to information asymmetry. Spence (1974) suggests that the willingness of a potential employee to incur the costs of education and training is a reliable signal of quality that can be used to overcome the information asymmetries in the labor market. In insurance markets, the inability of insurers to accurately identify the risk class of potential policyholders creates the potential for market failure. If premiums are set at a rate based on the risk of the pool of potential policyholders, the good risks will choose to forgo insurance and the pool of policies will make losses for the insurer. However, Rothschild and Stiglitz (1976) suggest that the willingness of good risks to forgo full insurance (e.g., through deductibles and coinsurance) may provide a reliable signal that will allow the insurer to charge them appropriate premium rates. The result is that there will be two kinds of policies in the market, partial insurance policies at lower rates for the good risks, and full insurance at higher rates for the higher risks.

In the context of financial markets, information asymmetry exists between managers and shareholders. In addition, incentive conflicts make managers' favorable public announcements regarding the firm's future prospects less credible. Therefore, it is necessary that managers signal information to investors in a credible way, e.g., their willingness to accept compensation in the form of stock options. The high degree of regulation of information in the financial markets is, in part, designed to reduce information asymmetry that could lead to reduced investor confidence and market failure.

Bibliography

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, **84**, 488–500.

Brigham, E. F. and Gapenski, L. C. (1997). *Intermediate Financial Management*, 5th edn. New York: International Thomson Publishing.

Rothschild, M. and Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *Quarterly Journal of Economics*, **90**, 629–49.

Spence, M. (1974). *Market Signaling*. Cambridge, MA: Harvard University Press.

Lerner index

Gilbert Becker

This is one of several measures developed by economists to establish the extent of MARKET POWER held by firms in a market. Lerner (1934) devised this measure, which is calculated as

$$L = \frac{(P - MC)}{P}$$

where P is the price of a firm's product and MC is the MARGINAL COST.

When price is equal to marginal cost, the lower limit value for L is zero. Larger values of L, indicating greater market power, have a theoretical upper limit of 1, if MC equals zero. Under PERFECT COMPETITION a firm's MARGINAL REVENUE (MR) from additional sales is constant and equal to price, which is determined by the market. Following the rule for PROFIT MAXIMIZATION, the firm finds the output where MR (and thus P) is equal to MC. As a result the Lerner index for such firms, which have no ability to influence the market, has a value of zero. For MONOPOLY and other imperfectly competitive markets, price typically is greater than marginal cost and thus the value of L is greater than zero. In general, the ability of a firm to set its price above marginal cost and sustain that differential over time is indicative of market power.

To the extent that MC approximates the firm's AVERAGE TOTAL COST, the Lerner index indicates the existence of positive ECONOMIC PROFIT resulting from this power. Although numerous empirical studies have used this method for several decades (see Church and Ware, 2000, for a survey), one difficulty with this index is that of measuring marginal cost. Often, to make the index operational, constant costs are assumed where marginal and average costs are equal. When this is done the value of the index must be examined over time in order to identify market power, since even perfectly competitive firms can earn positive economic profits in the short run.

It can also be shown that the Lerner index for a profit maximizing monopoly is inversely related to the price ELASTICITY of demand for the product. For example, a firm selling a product having an elasticity of demand equal to 3 has a Lerner index measure of market power of one third, or equivalently the ability to mark up its price by 33 percent above marginal cost. The more inelastic demand is, the greater the value of the index, indicating greater market power. Moreover, in OLIGOPOLY markets which follow a process of COURNOT COMPETI

TION, it can be shown that the Lerner index for any firm is positively related to its market share.

The Lerner index is also a measure of MARKET PERFORMANCE. The social loss arising from the allocative inefficiency of imperfectly competitive markets depends directly on the extent of the divergence between price and marginal cost in the market (see DEADWEIGHT LOSS).

Bibliography

Church, J. and Ware, R. (2000). *Industrial Organization: A Strategic Approach*. Boston: McGraw-Hill Irwin.
Lerner, A. (1934). The concept of monopoly and the measurement of monopoly power. *Review of Economic Studies*, **1**, 157 75.
Martin, S. (1994). *Industrial Economics: Economic Analysis and Public Policy*, 2nd edn. New York: Macmillan.
Maurice, S. and Thomas, C. (2002). *Managerial Economics*, 7th edn. Boston: McGraw-Hill Irwin.
Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

limit pricing

Robert E. McAuliffe

When a firm practices limit pricing it chooses its price (and therefore output) in such a way that the remaining market demand (RESIDUAL DEMAND) is insufficient for an entrant to cover its AVERAGE TOTAL COST. As originally described by Modigliani (1958), if an established firm were threatened by ENTRY in a single period, the limit price is the highest price it could charge without allowing entry to occur. Naturally, the limit price that would accomplish this objective depended on the expectations of potential entrants. Modigliani assumed that entrants would expect the established firm to continue producing at the entry limiting output even if entry occurred, equivalent to the assumption of COURNOT COMPETITION, where each firm believes its rivals will continue to produce at current levels. If there are ECONOMIES OF SCALE in production, limit pricing would allow established firms to earn economic profits while preventing entry, even when entrants have the same costs as incumbent firms (see ECONOMIC PROFIT). Gilbert (1989) refers to this pricing policy as "classic limit pricing."

Gaskins (1971) extended this model to con sider dynamic limit pricing where a firm is threatened by potential COMPETITION in the current period and in future periods, and where the rate of entry by new firms depended on the difference between the current price and their marginal costs. If the established firm set a high price it would earn high profits initially but would also encourage more entry and would decrease the price (and profits) in the future. A lower price set by the established firm would slow the rate of entry into the industry but would also lower profits. If the established firm had no cost or other advantages over entrants, ultimately it would lose its dominant position and the market would become perfectly competitive (*see* DOMINANT FIRM; PERFECT COMPETITION). But, as Jacquemin (1987) notes, if an existing firm has no cost advantages over entrants and is only able to use price as its strategic weapon, a competitive outcome is not surprising.

In the classic limit pricing model and the dynamic limit pricing model, potential compe tition constrains the MARKET POWER of existing firms who are forced to set prices below the MONOPOLY level. However, the as sumption about the entrant's expectations in these cases is restrictive. Why should an entrant expect the established firm to maintain its output level after entry has occurred if that is not the profit maximizing choice? In the post entry period both firms would earn higher profits by restricting output and raising price, so a threat by established firms to keep output constant after entry has occurred is not credible and an entrant should not believe it. For such a threat to be credible, the established firm must make commitments in the current period which make high output levels profitable in subsequent periods. Dixit (1980) showed that investments in EXCESS CAPACITY or in other SUNK COSTS could commit the established firm to maintain high output in subsequent periods. These sunk costs represent a signal to potential entrants that the established firm would respond aggressively to entry and make entry deterrence credible.

Limit pricing is an element of firm conduct in a market which, if successful, can affect MARKET STRUCTURE. In his survey of research in the area, Gilbert (1989) found weak evidence in support of the classic limit pricing theory and the dynamic limit pricing theory. Established firms do not appear to set prices at the monopoly level, nor do they set prices so low that entry is deterred (see Masson and Shaanan, 1982; Greer, 1992). However, in his survey of executives, Smiley (1988) found that managers of major American companies preferred higher ADVER TISING expenditures and PRODUCT PROLIF ERATION as strategic weapons to deter entry and were less likely to rely on excess capacity and limit pricing.

See also *accommodation; credible strategies; profit maximization; signaling*

Bibliography

Dixit, A. (1980). The role of investment in entry deter rence. *Economic Journal*, **90**, 95 106.

Encaoua, D., Geroski, P., and Jacquemin, A. (1986). Strategic competition and the persistence of dominant firms: A survey. In J. E. Stiglitz and G. F. Mathewson (eds.), *New Developments in the Analysis of Market Structure*. Cambridge, MA: MIT Press.

Gaskins, D. (1971). Dynamic limit pricing: Optimal pricing under threat of entry. *Journal of Economic Theory*, **2**, 306 22.

Gilbert, R. J. (1989). The role of potential competition in industrial organization. *Journal of Economic Perspec tives*, **3**, 107 28.

Greer, D. F. (1992). *Industrial Organization and Public Policy*, 3rd edn. New York: Macmillan.

Jacquemin, A. (1987). *The New Industrial Organization: Market Forces and Strategic Behavior*. Cambridge, MA: MIT Press.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Masson, R. and Shaanan, J. (1982). Stochastic dynamic limit pricing: An empirical test. *Review Of Economics and Statistics*, **64**, 413 23.

Modigliani, F. (1958). New elements on the oligopoly front. *Journal of Political Economy*, **66**, 215 32.

Smiley, R. (1988). Empirical evidence on strategic entry deterrence. *International Journal of Industrial Organ ization*, **6**, 167 80.

linear regression

Robert E. McAuliffe

Linear regression allows managers and research ers to estimate (or quantify) how much a

particular variable is affected by another and to determine whether the relationship is statistic ally significant. Although nonlinear models and techniques are also available (see Greene, 2002), a wide class of models can be estimated with linear regression methods, which may ex plain their popularity. Even so, linear regression techniques are appropriate only when the *true* underlying relationship between the dependent variable and the independent variable(s) is in fact linear or can be transformed into a linear relationship. If the model is misspecified, then the resulting estimates are likely to be biased.

In a linear regression, also referred to as **ordi nary least squares**, the true population rela tionship between the dependent variable, Y_i, and the independent variable, X_i, is assumed to be linear. For example, in the simplest case with only one independent variable, the relation to be estimated is:

$$Y_i = \beta_0 + \beta_1 \times X_i + v_i \qquad (1)$$

where the Y_i are observations on the *dependent variable*, the X_i are observations on the *independ ent* or *explanatory variable*, and the v_i are the error terms. The coefficients β_0 and β_1 are the population parameters to be estimated and re flect the effect of the variable X on Y. For example, a manager or researcher might be inter ested in the effect of ADVERTISING on sales, and the expected relationship might be given as in (1) above. The least squares estimators of β_0 and β_1 are obtained by minimizing the sum of the square of the estimated sample errors. If we let $\hat{\beta}_0$ and $\hat{\beta}_1$ represent the least squares estima tors of β_0 and β_1 in (1) above, then the estimated sample error, $\hat{v}_i$, can be defined as

$$\hat{v}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 \times X_i \qquad (2)$$

Given equation (2), the least squares estimators are obtained by minimizing the sum of the squared sample errors below with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$. Thus we minimize

$$\sum_{i\,1}^{n} \hat{v}_i^2 \qquad (3)$$

by choosing the best values for $\hat{\beta}_0$ and $\hat{\beta}_1$ where the summation is over all n observations. It can

be shown that the least squares estimators, $\hat{\beta}_0$ and $\hat{\beta}_1$, are given by:

$$\hat{\beta}_1 = \frac{\sum (Y_i - \bar{Y}) \times (X_i - \bar{X})}{\sum (X_i - \bar{X})^2}$$
$$= \frac{\text{cov}(Y, X)}{\text{var}(X)} \qquad (4)$$
$$\text{and } \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \times \bar{X}$$

where $\text{cov}(Y,X)$ is the COVARIANCE of the de pendent variable, Y, with the independent variable, X; $\text{var}(X)$ is the VARIANCE of the inde pendent variable; and $\bar{Y}$ and $\bar{X}$ are the sample means of Y and X. Therefore, if the goal is to choose linear estimators to make the sum of the squared estimated errors as small as possible, the estimators in equation (4) above will satisfy this requirement. In fact, it can be shown that the ordinary least squares estimators are the *best* in the class of unbiased linear estimators (see Maddala, 2001; Greene, 2002) if the assumptions regarding the error terms hold (*see* ERROR TERMS IN REGRESSION).

It can be shown that the least squares estima tors in equation (4) above have expected values equal to the population parameters when the assumptions regarding the error terms in linear regression hold. The standard error of these estimators is calculated by most statistical pro grams (*see* STANDARD ERROR OF THE COEF FICIENT; STANDARD ERROR OF ESTIMATE), and CONFIDENCE INTERVALS and t statistics can then be calculated to determine whether the independent variable significantly affects the de pendent variable (*see* T STATISTIC). One meas ure of how well the estimated model fits the data is the coefficient of determination, or R^2, which measures the percentage of the variation in the dependent variable explained by the independ ent variable. A high value of the R^2 (i.e., a value close to 1) indicates that most of the variation in the dependent variable is explained by the inde pendent variable.

In many cases the underlying relationship be tween the dependent and independent variables is nonlinear but can be transformed into a linear relation. For example, the COBB–DOUGLAS PRODUCTION FUNCTION is:

$$q = AL^{\alpha}K^{\beta} \qquad (5)$$

where q is the output produced and L and K are the labor and capital inputs, respectively. This production function can be transformed by taking logarithms of both sides to obtain:

$$\mathbf{log}(q) = \mathbf{log}(A) + \alpha \times \mathbf{log}(L) + \beta \times \mathbf{log}(K) \tag{6}$$

which is now a linear relationship and can be estimated by least squares. A convenient feature of a logarithmic regression as above is that the estimated coefficients are elasticities. The coefficient α above is the ELASTICITY of output with respect to a change in labor input, while β is the elasticity of output with respect to a change in capital input.

Another possibility when the underlying relationship is nonlinear is to create additional explanatory variables which are simply higher powers of X_i:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 \times X_i + \hat{\beta}_2 \times (X_i)^2 + \hat{\beta}_3 \times (X_i)^3$$

This transformation allows a researcher to estimate a cubic relationship between the independent variable and the dependent variable, although MULTICOLLINEARITY might occur.

When there are several explanatory variables in the regression equation (multiple regression), the basic features of the least squares estimators remain but some additional problems are created. First, the estimated coefficient in a multiple regression represents the additional or incremental effect of the independent variable, *holding the other independent variables constant*. The multiple regression would be written as:

$$q_i = \hat{\beta}_0 + \hat{\beta}_1 \times X_{1,i} + \hat{\beta}_2 \times X_{2,i} + \hat{\beta}_3 \times X_{3,i} + v_i$$

where the $X_{k,i}$ represent the i observations on the different explanatory variables. For example, the Cobb–Douglas production function might specify $X_{1,i}$ as observations of the logarithm of units of labor employed to produce output, $X_{2,i}$ as the logarithm of units of capital employed, and $X_{3,i}$ as the logarithm of units of some other input, such as energy or materials. The coefficient $\hat{\beta}_0$ would then be an estimate of the logarithm of technical change, while the coefficient $\hat{\beta}_1$ would be the estimated elasticity of output given a change in labor input, holding all other variables constant.

A problem in multiple regression is multicollinearity, where there is correlation between the explanatory variables. When one of the independent variables is correlated with one or more of the remaining explanatory variables, the data make it difficult to determine the independent contributions of each of the X_k variables. As a result, the estimated standard error of the coefficients will be larger, and this will make it appear as though one or more of the independent variables have no significant effect on the dependent variable when in fact they may.

When estimating a multiple regression, it is important to avoid specification errors in the model. A researcher may, for example, include irrelevant explanatory variables in the regression or may omit variables that should be included. Generally, including irrelevant explanatory variables will not bias the estimated coefficients, but omitting variables that should be included in the regression will cause the least squares estimators to be biased (see Maddala, 2001; Greene, 2002). One indication that relevant variables may have been omitted from the regression is the presence of AUTOCORRELATION in the estimated residuals.

Bibliography

Greene, W. H. (2002). *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.

Gujarati, D. N. (2002). *Basic Econometrics*, 4th edn. New York: McGraw-Hill Irwin.

Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.

lock-in

Robert E. McAuliffe

This term became very important in the Internet literature and was thought to play a role in network effects (*see* NETWORK EXTERNALITIES). It begins with the concept of SWITCHING COSTS. Lock in occurs when the switching costs of using a new technology or changing to a different technology are so great that users do not switch despite advantages of the new technology. It was thought that consumers who utilized the worldwide web would incur lock in because they were required to become familiar

with the websites they used. The argument became acceptable almost immediately and was not given a great deal of attention (see Arthur, 1989).

Shapiro and Varian (1999) viewed lock in as a source of MARKET POWER that firms could exploit. They suggested that a firm could use lock in to gain an advantage over con sumers or rival firms, perhaps by giving use of a free but limited version of software but charging customers a fee for a more attractive and capable package. The assertion was that the free version of the software would establish market share for the company, while sales of the more able package would generate revenues for the firm.

Problems arise, however, when it becomes necessary to identify the specific characteristics of a product that create a lock in effect. Leibow itz and Margolis (1990, 1995) explored this notion and found there were no examples of strong lock in in US history. They found that in key instances of supposed lock in, mainly for the QWERTY keyboard layout (see David, 1985) and Beta versus VHS, the market deter mined outcome was the best. For example, the QWERTY keyboard configuration was not found to be worse than the Dvorak keyboard; in fact, a GSA study in the 1950s found that the QWERTY keyboard was comparable. Further more, although the Beta format was available much earlier, VHS was superior because it offered a longer recording time. These research ers concluded that if a company has a choice between getting to market first and developing the product well, the company would be best off if it pursued the latter option.

Bibliography

Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *Economic Journal*, **99**, 116 31.

David, P. A. (1985). CLIO and the economics of QWERTY. *American Economic Review*, **75**, 332 7.

Leibowitz, S. J. (2002). *Rethinking the Network Economy*. New York: American Management Association.

Leibowitz, S. J. and Margolis, S. E. (1990). The fable of the keys. *Journal of Law and Economics*, **33**, 1 26.

Leibowitz, S. J. and Margolis, S. E. (1995). Path depend ence, lock-in, and history. *Journal of Law, Economics and Organization*, **11**, 205 26.

Shapiro, C. and Varian, H. R. (1999). *Information Rules*. Boston: Harvard Business School Press.

long run cost curves

Robert E. McAuliffe

For the firm, the long run refers to the length of time required so that all inputs in production are variable. There are no FIXED COSTS in the long run. Since the firm is free to choose its capacity level in the long run, this is also considered the planning horizon. The long run is not fixed in calendar time but may vary across industries and even among firms within the same industry if their contractual commitments differ. From the perspective of the industry, Stigler (1966) sug gested that the long run was the time required for a market or industry to fully adjust to "new conditions," and the period of adjustment re quired would depend upon the questions under consideration.

The long run average cost (LRAC) curve is the planning curve for the firm because it shows the minimum average cost of production using plants of varying sizes. As such it envelops the short run average cost (SRAC) curves for differ ent capacity levels and is typically assumed to be U shaped. Associated with the LRAC curve is a long run marginal cost (LRMC) curve, which lies below the LRAC curve when the LRAC is falling (due to ECONOMIES OF SCALE), inter sects the LRAC curve at its minimum, and lies above the LRAC curve when the LRAC curve is rising (when DISECONOMIES OF SCALE occur), as in figure 1.

The LRMC curve reflects the change in TOTAL COSTS when an additional unit of output is produced, given that all inputs are adjusted optimally (including CAPITAL). When capital expenditures are discrete and cannot be adjusted to produce one additional unit, such as when the scale of the plant is changed, the LRMC curve shows the changes in total costs moving to the next scale of oper ation (see Sexton, 1995). In these cases, the LRAC curve will not be as smooth as in figure 1 and will follow the individual short run average cost curves more closely. If a manager antici pated sales of 1 million units per period, then
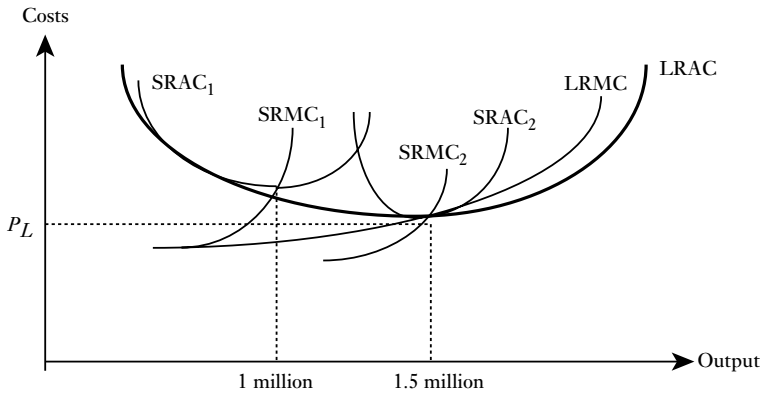
Figure 1   Long run and short run cost curves

she would choose capacity level SRAC$_1$ above. However, if sales were expected to be 1.5 million units per period, SRAC$_2$ would be the best capacity choice and, in this example, it is also the MINIMUM EFFICIENT SCALE. For this plant, both the LRAC and SRAC curves are at their minimum points at 1.5 million units of output per period, and so the LRMC curve intersects the short run marginal cost (SRMC$_2$) curve at this point as well. It should also be noted that for each point on the LRMC curve a SRMC curve passes through it. Since some factors of production are fixed in the short run, the SRMC curve should be steeper than the LRMC curve where all factors of production can be adjusted optimally (Friedman, 1976; Shughart, Chappell, and Cottle, 1994; *see* SHORT RUN COST CURVES).

The output level of 1.5 million units at a price P_L represents the long run EQUILIBRIUM for a firm in a competitive industry, where the representative firm earns zero

ECONOMIC PROFIT and operates a plant at the minimum efficient scale. At this price there will be no net ENTRY into or net EXIT from the industry unless the conditions of supply or demand change.

Bibliography

Friedman, M. (1976). *Price Theory*, 2nd edn. Chicago: Aldine.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

Sexton, R. L. (1995). *Microeconomics*. Englewood Cliffs, NJ: Prentice-Hall.

Shughart, W. F., Chappell, W. F., and Cottle, R. L. (1994). *Modern Managerial Economics*. Cincinnati, OH: South-Western Publishing.

Stigler, G. J. (1966). *The Theory of Price*, 3rd edn. New York: Macmillan.

M

make or buy decisions

Robert E. McAuliffe

A firm's decision to make an input in produc
tion rather than buy that input in the market is
an important one that firms frequently fail to
reconsider. What a firm chooses to produce is
one of the most fundamental decisions it can
make. But if markets are so efficient in provid
ing resources, why should firms even exist?
Coase (1937) answered that there were costs
from using the market to obtain resources
just as there are costs of producing them within
the firm. The TRANSACTIONS COSTS of using
the market must then be weighed against the
coordination costs of producing the input in
ternally.

 Williamson (1985) argued that internal pro
duction requires additional bureaucracy to
manage the new activity along with higher pro
duction costs. But if the firm obtained the input
in the market, it would be costly (in terms of
time) to locate suitable suppliers, monitor qual
ity levels, and insure that supplies will be de
livered on time. When the part is unique and
very specialized, it is costly to use the market
because there may only be one or two suppliers.
In this case, the firm may decide to produce the
part internally.

 Porter (1980) suggests that VERTICAL INTE
GRATION may also achieve strategic goals for
the firm. For example, a firm may want to im
prove the quality or lower the production costs
of its product but would need closer coordin
ation with its suppliers to accomplish this goal.
Another possibility is that internal production
will improve the firm's flexibility to produce
new products.

Bibliography

Coase, R. H. (1937). The nature of the firm. *Economica*, **4**, 386 405.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.

Shughart, W. F. (1990). *The Organization of Industry*. Boston: Richard D. Irwin.

Williamson, O. E. (1985). *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. New York: Free Press.

marginal cost

Robert E. McAuliffe

The marginal cost is the change in TOTAL
COSTS due to a unit (or incremental) change in
output. That is, for discrete changes in output,
marginal cost is given by:

$$MC = \frac{\Delta TC}{\Delta q} \qquad (1)$$

where MC is the marginal cost, ΔTC is the
change in total costs, and Δq is the change in
output. In the SHORT RUN, FIXED COSTS do
not vary with output, so marginal costs can also
be written as:

$$MC = \frac{\Delta TVC}{\Delta q} \qquad (2)$$

where ΔTVC is the change in TOTAL VARI ABLE COST. If the total cost function is known or has been estimated, marginal cost would be the derivative of total costs with respect to output. Therefore, if total costs are:

$$\text{TC} = 1,000 + 5 \times q + 8 \times q^2 \qquad (3)$$

marginal costs would be:

$$\text{MC} = \frac{d\text{TC}}{dq} = 5 + 16 \times q \qquad (4)$$

Understanding marginal costs is crucial for op timal decision making because, unless there are resource constraints, any activity should be con tinued as long as the additional (or marginal) benefits exceed the marginal costs. If two or more activities have marginal benefits that exceed their marginal costs but a manager cannot fully fund all of them, then each activity should be funded until each provides the same marginal benefits per dollar (see PROFIT MAXIMIZA TION; UTILITY MAXIMIZATION). For any *change* in an activity (production, pricing, AD VERTISING, etc.), a manager must compare the marginal benefits of the change against the mar ginal costs to make the best decision.

For example, consider a manager whose only variable input is labor. The change in total costs will then be the change in labor input (ΔL) multiplied by the wage rate (w), and since mar ginal costs are the change in total variable costs divided by the change in quantity (Δq), from equation (2) we have:

$$\text{MC} = \frac{\Delta\text{TVC}}{\Delta q} = \frac{w \times \Delta L}{\Delta q} = w \times \left\{\frac{\Delta L}{\Delta q}\right\} \quad (5)$$

The term in braces is the inverse of MARGINAL PRODUCT of labor, so equation (5) shows the relationship between marginal costs, variable input costs (the wage rate, w), and productivity (the marginal product of labor). An increase in the cost of variable inputs, w, will increase the marginal cost of production, while an increase in productivity (a rise in the marginal product of labor) will reduce marginal costs, as would be expected (see Shughart, Chappell, and Cottle, 1994; Friedman, 1976).

As Baumol (1977) notes, managers frequently use average values rather than marginal values when making decisions, and this can lead to incorrect conclusions. This occurs because much of the accounting information provided to managers is in the form of average of total figures. In addition, marginal calculations may require information a company does not yet have, such as the marginal benefits and costs of an increase in advertising. The important issue is not whether past advertising expenditures have generated net revenues that exceeded the costs, but whether any increase (or decrease) in adver tising is justified. Finally, because marginal costs reflect the change in total costs for the last unit(s) produced, they reflect changes more quickly than average total costs, which are based on the costs for all units produced to that point (see AVERAGE TOTAL COST). Thus even if the aver age increase in net revenues from advertising fell by a small amount, the *marginal* change could be substantial and profits could be increased by reducing advertising expenditures.

Bibliography

Baumol, W. J. (1977). *Economic Theory and Operations Analysis*. Englewood Cliffs, NJ: Prentice-Hall.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Friedman, M. (1976). *Price Theory*, 2nd edn. Chicago: Aldine.

Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Pub-lishing.

Shughart, W. F., Chappell, W. F., and Cottle, R. L. (1994). *Modern Managerial Economics*. Cincinnati, OH: South-Western Publishing.

marginal product

Lidija Polutnik

The marginal product of any input, such as labor or capital, measures the change in TOTAL PRODUCT as a result of a small change in the usage of that input, holding all other inputs constant. For example, the marginal product of labor or capital is the additional output from using one more labor or machine hour.

The marginal product of an input that is infi nitely divisible can be obtained by taking the

partial derivative of the production function with respect to that input (*see* PRODUCTION FUNCTIONS). In the relevant range of production, marginal product is positive. An increase in the marginal product (or an increase in productivity) of an input means that through better technology, more efficient management, or an increase in work effort, the company is able to obtain more output from the same amount of inputs. Alternatively, marginal product can be improved through the restructuring of the company by decreasing the number of work hours while holding the addition to total output constant.

The total product and marginal product curves demonstrate the LAW OF VARIABLE PROPORTIONS. This law states that as the quantity of a variable input increases, holding the amount of other productive factors constant, beyond some point the marginal product of the variable input begins to decrease; total product continues to increase, though, at a decreasing rate. DIMINISHING RETURNS arise because the fixed amount of plant and equipment is gradually spread among an even greater number of workers, leaving a smaller amount of CAPITAL to work with for each variable input.

By enabling us to measure the change in total product as a result of a change in a variable input, marginal product permits comparisons with the short run cost of production (*see* SHORT RUN COST CURVES). If the costs of labor and machine hours are known, we can calculate the MARGINAL COST of production by dividing an input's cost by the marginal product of the input. Therefore, marginal costs are inversely related to the marginal product of the variable inputs. For example, when the marginal product of labor is falling, the marginal cost of output is increasing, and vice versa. When the marginal product of labor is constant, additional units can be produced with constant marginal cost.

Knowledge of the marginal products of inputs enables managers to determine the short run optimal level of employment of capital and labor in the production process. In order to maximize total profit (*see* PROFIT MAXIMIZATION), inputs should be hired up to the point where the marginal input cost equals the MARGINAL REVENUE product of labor. Marginal input cost is defined as the amount that an add itional unit of the variable input adds to TOTAL COSTS. Marginal revenue product is defined as the amount that an additional unit of the variable unit adds to TOTAL REVENUE and equals the marginal product of the variable input times the marginal revenue resulting from the increase in output produced.

Bibliography

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Publishing.

marginal rate of substitution

Gilbert Becker

The marginal rate of substitution of good X for good Y (MRS$_{x,y}$) is defined as the amount of good Y which the consumer is willing to give up in exchange for one additional unit of good X, in order to maintain the same level of TOTAL UTILITY. Since the same total utility is achieved from the two choices, MRS$_{x,y}$ concerns the movement along a consumer's indifference curve (*see* INDIFFERENCE CURVES). Specifically, it measures the slope of the indifference curve between any two points. This identifies the trade off of one good for another that the consumer is *willing* to make, independent of the prices of the two goods which indicate the trade off conditions *required* by the market.

The marginal rate of substitution between two goods describes the relative importance of the two goods to the consumer and it is defined as the ratio of the marginal utilities of the two goods. For example, a MRS$_{x,y}$ with the value of 3 to 1 indicates that the consumer is currently willing to give up 3 units of good Y in order to receive 1 extra unit of good X. Here, the MARGINAL UTILITY of good X(MU$_x$) is three times that of good Y. This can be formally expressed as:

$$\text{MRS}_{x,y} = -\frac{\Delta Y}{\Delta X} = \frac{\text{MU}_x}{\text{MU}_y} = 3$$

where Δ indicates the change in the quantity of the good in question.

Since the marginal utility of either good will typically decrease with greater consumption of that good, the marginal rate of substitution and therefore the slope of the indifference curve is typically not constant. One exception is the case where the two goods in question are perfect SUBSTITUTES. Here the marginal rate of substitution will be constant, e.g., 2 aspirin may be substituted for every 1 acetaminophen tablet.

The rate at which consumers are willing to trade off one product for another can be altered by changing their tastes. Here, ADVERTISING and other forms of promotion can be useful.

See also *diminishing marginal utility; utility maximization*

Bibliography

Maurice, S. and Thomas, C. (2002). *Managerial Economics*, 7th edn. Boston: McGraw-Hill Irwin.

Nicholson, W. (2000). *Intermediate Microeconomics and its Application*, 8th edn. New York: Harcourt.

marginal rate of technical substitution

Gilbert Becker

The production of virtually every good requires the use of two or more inputs such as labor, capital, and raw materials. Typically, these inputs can be used in varying proportions to generate the same level of output. For example, in agriculture varying amounts of labor and capital can be used to plant or harvest the same number of acres of land per day. Modern farming techniques in the US and other developed nations use large harvesting combines requiring relatively few workers, while less sophisticated farming operations often use a highly labor intensive approach. Since reducing the amount of any one input generally tends to reduce total output, a greater quantity of a second input is needed to compensate for the lost output. The marginal rate of technical substitution measures the rate at which any one input must be substituted for one unit of another input in order to maintain a constant level of output.

Formally, the production of varying levels of output using different combinations of inputs is described using an isoquant map (*see* ISO QUANT ISOCOST CURVES). The marginal rate of technical substitution is measured by the slope of a given isoquant curve between any two points on that curve. Since increased use of any one input tends to reduce that input's marginal productivity, the marginal rate of technical substitution tends to decrease with greater substitution toward an input. That is, the isoquant curve tends to be convex.

The marginal rate of technical substitution between inputs is important information for managers selecting the optimal combination of inputs. Just as the consumer trade off between two goods is determined by the ratio of the MARGINAL UTILITY of the two *goods* (*see* MARGINAL RATE OF SUBSTITUTION), the marginal rate of technical substitution is determined by the ratio of the MARGINAL PRODUCT of the two *inputs*. Cost minimization requires that the marginal rate of technical substitution between inputs be equal to the ratio of the prices of those inputs. This is the production requirement equivalent to the consumer optimization problem (*see* UTILITY MAXIMIZATION).

Bibliography

Browning, E. and Zupan, M. (2004). *Microeconomics: Theory and Applications*, 8th edn. New York: John Wiley.

Maurice, S. and Thomas, C. (2002). *Managerial Economics*, 7th edn. Boston: McGraw-Hill Irwin.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

marginal revenue

Gilbert Becker

Marginal revenue is the change in the TOTAL REVENUE resulting from a one unit increase in the sales of a product. It is important to a firm because PROFIT MAXIMIZATION requires the firm's managers to find the output level where this extra revenue equals the MARGINAL COST. While greater sales resulting from price reductions generally cause marginal revenue to fall, marginal cost eventually rises with greater

output. Thus, any output below the level where the two are equal leaves the possibility for additional profits to be made, but output beyond that level costs more to produce than the gain in revenues from the additional sale.

The nature of the relation between marginal revenue and the level of output sold varies with the basic type of market under consideration (*see* BASIC MARKET STRUCTURES). Under PERFECT COMPETITION the firm, being too small to influence market outcomes, can sell as much output as it desires without changing the market price. As such, for any additional sales, the marginal revenue is constant at the level of the market generated price. In other words, the firm faces a demand which is perfectly elastic for its own product (*see* ELASTICITY).

Under OLIGOPOLY and other forms of imperfect competition, additional sales may occur as a result of a change in the quality of the product or a change in ADVERTISING (*see* DORFMAN–STEINER CONDITION), or they may be brought about by a change in the product's price. The LAW OF DEMAND indicates that, all else being equal, an increase in the number of units sold requires a reduction in price. In the absence of PRICE DISCRIMINATION, the marginal revenue from an additional unit sold declines as the lower price needed to generate a larger number of unit sales will offset the gains from the increased quantity sold. Moreover, the estimation of marginal revenue becomes more difficult in oligopoly since the amount of additional sales resulting from a price reduction depends in part on the reaction of rivals, who may or may not respond with price cuts of their own.

Although it is not always consistent with profit maximization, managers may at times wish to maximize sales revenues (*see* OBJECTIVE OF THE FIRM). Here, marginal revenue and its relation to price elasticity of demand are important. Where demand is elastic, a decline in price results in a large increase in unit sales, and total revenue is growing. In this case, the marginal revenue from the additional units sold is positive. For revenue maximization, production should continue to increase until marginal revenue is zero, i.e., where demand is unit elastic.

See also perfect competition; profit maximization

Bibliography

Maurice, S. and Thomas, C. (2002). *Managerial Economics*, 7th edn. Boston: McGraw-Hill Irwin.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

Schiller, B. (2003). *The Micro Economy Today*, 9th edn. New York: Irwin McGraw-Hill.

marginal utility

Gilbert Becker

Marginal utility is defined as the change in the TOTAL UTILITY resulting from the consumption of one additional unit of a good by a consumer. For any individual good, consumer behavior typically follows the law of DIMINISHING MARGINAL UTILITY, which states that in any given time period, as the rate of consumption of a good rises, the additional utility or satisfaction acquired by the consumer eventually declines.

It can be shown that for an individual who is purchasing several goods simultaneously, the maximization of total utility requires equating the marginal utility per dollar spent on each good. More formally, this ideal combination of goods occurs when

$$MU_1/P_1 = MU_2/P_2 = \ldots = MU_i/P_i \quad (1)$$

where MU and P indicate the marginal utility and price of the ith good. An imbalance between any two of these ratios, for example

$$MU_2/P_2 > MU_1/P_1 \quad (2)$$

would result in a situation where a reallocation of the consumer's income toward purchasing more of good 2 would cause a net increase in total utility. This is because the additional satisfaction gained from consumption of an extra unit of good 2 would more than offset the loss of utility from the decreased consumption of good 1. An inequality such as in equation (2) would occur if, for example, the price of the second good, P_2, was decreased. The equation demonstrates why a decrease in the price of this good would lead to a shifting away from good 1 and an increase in the purchase of good 2 (*see* SUBSTITUTION EFFECT). This helps to generate the negative

relationship between the price of good 2 and its quantity demanded, which is the foundation of the LAW OF DEMAND. Similarly, equation (2) can be used to demonstrate how a manager may create a more favorable position for a product by using increased ADVERTISING. This strategy, when successful, creates imbalances between the ratios by altering the marginal utility of one good relative to the others.

The marginal utility of income to a gambler, or profit to an investor, can be used to identify the three categories of risk behavior which individuals may portray. An individual is identified as a risk preferrer if his marginal utility of income grows at an increased rate as his level of income rises. An individual is said to show RISK AVERSION if the marginal utility of an extra dollar of income grows at a decreasing rate as income rises, and RISK NEUTRALITY if the marginal utility of an extra dollar of income is constant (see Varian, 1992). A risk averter typically sees risk as being undesirable, and is willing to tolerate greater uncertainty in an outcome only if rewarded with a greater expected return. A risk lover is a gambler who prefers the increased uncertainty (*see* EXPECTED PRESENT VALUE).

See also *marginal rate of substitution*

Bibliography

Maurice, S. and Thomas, C. (2002). *Managerial Economics*, 7th edn. Boston: McGraw-Hill Irwin.
Nicholson, W. (2000). *Intermediate Microeconomics and its Application*, 8th edn. New York: Harcourt.
Schiller, B. (2003). *The Micro Economy Today*, 9th edn. New York: Irwin McGraw-Hill.
Varian, H. R. (1992). *Microeconomic Analysis*, 3rd edn. New York: Norton.

market definition

Gilbert Becker

An economic definition of the term "market" is the group of all firms willing and able to sell a similar product or service to the same potential buyers. As such, the market definition for a given product is the identification of the relevant group of sellers who may be seen as being in COMPETITION with one another for the sale of that product. Two boundaries are used to define a market:

1 the product boundary;
2 the geographic boundary.

Properly identifying these boundaries for a particular market will result in a more accurate count of the number and size distribution of firms in that market. This information is useful for a firm in developing its competitive strategy. For markets under PERFECT COMPETITION the product boundary is relatively easy to establish since the product is homogeneous across firms. Given the complexity of most markets though, and especially those where PRODUCT DIFFERENTIATION exists, this boundary becomes blurred. Similar difficulties arise in identifying correct geographic boundaries. Thus, managers should be aware that the term market may be a source of significant confusion. Sales representatives and other subordinates may erroneously use the term to describe their personal selling area or limit their definition to the specific product line associated with their selling effort. These approaches may underestimate the true scope of the economic market in question.

Market definition is also essential for the government and the courts in developing antitrust policy and in deliberating antitrust cases. For example, the CLAYTON ACT proscribes mergers which substantially lessen competition "in any line of commerce" (i.e., product market) and "in any section of the country" (i.e., GEOGRAPHIC MARKET). Thus the proper identification of market boundaries will be critical in indicating the degree to which competition may have declined. An expanded market definition, for example, which includes a greater number of rivals, will naturally tend to diminish the market share of the merging firms and thus make the merger less potentially onerous in the eyes of the antitrust enforcers. Moreover, in cases of monopolization, proper identification of the market is essential to determine the existence of MARKET POWER, which is the first necessary requirement for a finding of an antitrust violation (*see* MICROSOFT ANTITRUST CASE; SHERMAN ACT).

The central issues in market definition are where and how these boundaries are to be drawn. Numerous authors including Church and Ware (2000) have recognized that market definition and measurement, when applied to antitrust policy, may differ from that of the simple economic definition given above. As such, several tools have been introduced and are categorized below as either theoretical or practical measures, each having its own relative strength in terms of its use in purely economic versus public policy applications.

THEORETICAL MEASURES

CROSS ELASTICITIES are one set of market measurement tools which have been used for decades. The cross elasticity of demand between two goods measures the percentage change in unit sales of one firm's product given a 1 percent change in the price of a rival's product. As such it measures the degree to which the two goods are SUBSTITUTES in the eyes of consumers. A high cross elasticity measure indicates close substitutability and thus the firms in question should be counted as being in the same market. Similarly, cross elasticities of supply may be considered. High values for this measure indicate a firm's willingness and ability to produce more of their own good in reaction to a rival's price increase. It may also indicate a similarity in production technological processes between firms and the potential to switch product lines and readily and easily enter a product market where price has been increased above the competitive level. As a result, both measures of cross elasticities are needed to properly define the relevant product market. These measures attempt to determine an economic definition of a market by examining relevant rivals who establish a market's price.

While these measures are conceptually appealing, Scherer and Ross (1990), Shepherd and Shepherd (2004), and others have identified several difficulties that arise in their application. First, what constitutes a "high" value for cross elasticity is necessarily an arbitrary decision. Moreover, possible values for these measures fall along a continuum. Distinct gaps between high and low values which could indicate a break in the substitutability continuum for these goods may not arise in the actual estimates examined.

Third, two different estimates for the cross elasticity of demand between two goods A and B will be generated for the relatively similar cases of a 1 percent price increase in product A, and a 1 percent price decrease in product B.

Finally, the correct estimation of cross elasticities may require using price changes originating from the competitive price level, which may not be readily available. Where market power already exists and a firm's price is set significantly above that of its rivals, its cross price elasticity of demand may be higher since further price increases may induce considerably more consumer reaction than before. A small price increase for product A may induce a larger change in product B's sales when A's price is 50 percent above B's than when it is only 5 percent above that of B. Thus, if a monopolist already using its market power has set its price above the competitive level, the cross elasticity measured at that price level may be high, indicating the existence of many rivals and a broad market. The high cross elasticity measure here may lead to the erroneous conclusion that market power is limited, when in fact that power is already in use.

As a result, more recent antitrust analysis has focused on defining a market by examining the existence of sustainable market power. Perhaps the first such measure is the LERNER INDEX, which examines the extent by which a firm's price exceeds its MARGINAL COST. Since the Lerner index can be shown to be inversely related to a product's own price ELASTICITY of demand, the latter may also be used to help identify an antitrust market.

Boyer (1979) developed the basis of a theoretically more advanced measure which has been used by the federal antitrust enforcement agencies. This method can be used for both product market and geographic market definitions. The method involves examining a hypothetical monopolist in the geographic area in question, and its ability to introduce a small but significant nontransitory increase in prices (SSNIP) above the current level. Boyer argues that a market can be identified, from the point of view of a single firm, as the smallest group of rivals necessary to organize in order to successfully act as a cartel by imposing such a SSNIP (see CARTELS). If this group of firms (or hypothetical monopolist)

could raise prices without bringing about reaction by other firms that was sufficient to force a retreat in the cartel price hike, then these other firms can correctly be seen as being outside of the market. This may be due to their producing a sufficiently different (nonsubstitutable) product, or because their business takes place in a geographically distinct location, or even simply because their reaction was slow or limited. If, on the other hand, the initial group of firms in question could not successfully collude for a significant period of time without feeling the disciplinary effects of the other rivals' increases in supply, then the relevant market should be expanded to include these additional firms. In revising its guidelines for mergers in 1982 and 1992, the Department of Justice (DOJ) introduced a variation of Boyer's method, using a 5 percent price rise as the SSNIP to be examined.

Several shortcomings of this new method have been cited. Boyer's work recognized that problems similar to those involved with using cross elasticities still exist with the new approach. Moreover, Scherer and Ross argue that the use of the 5 percent rule in a situation in which market power already exists will tend to yield biased market definitions which could increase the likelihood of allowing the mergers that are being examined. Shepherd and Shepherd (2004) argue that estimates concerning future responses to a hypothetical monopolist's actions are speculative, and that what constitutes a sufficient reaction to a "small" and "significant" price rise is arbitrary. Finally, they maintain that both this measure and that of cross elasticity of supply improperly treat the question of BARRIERS TO ENTRY by combining it with the definition of market, thereby blurring the distinction between potential rivals and those actually in existence.

PRACTICAL MEASURES

Perhaps because of the weaknesses cited above, SSNIP has not yet been widely accepted by the courts. Instead, courts continue to use more pragmatic indicators. Given the complexity of markets, it remains unlikely that any one measure can accurately and completely determine the boundaries of markets. Perhaps realizing this,

the US Supreme Court has offered its own guides. (For an expansive historical examination of the US courts' development of the concept of market definition, see Werden, 1992.) In an important antitrust case involving the Du Pont Company's monopolization of the cellophane market, the court established the standard of "reasonable interchangeability" of products as the measure for market delineation (see Breit and Elzinga, 1996: 196–9 for details). It cited a product's price, qualities, and uses as being significant in examining this demand side substitutability. Shortly thereafter, in 1962, Chief Justice Warren, in a merger case involving the Brown Shoe Company, identified several additional market boundary determinants, including public recognition of a market, unique production facilities, differing prices among rivals, and even specialized vendors (Breit and Elzinga, 1996: 133–41). Moreover, the court here recognized that the Clayton Act allowed for examining submarkets for antitrust purposes.

Shepherd and Shepherd (2004) include, among other items, transportation costs relative to the product's value and shipping distances. These often limit competition to a certain geographic area. In many antitrust cases, the type of actual shipping data needed is readily available as evidence. Moreover, Breit and Elzinga's 20 percent rule gives a simple tool for discerning those who influence the price of a product and should be considered as market participants. Finally, numerous other forms of qualitative data, from surveys of consumer behavior to information regarding SEARCH COSTS, have been considered.

This list of market boundary variables is almost certainly overinclusive for any one case, but some items such as shipping information clearly warrant examination in many cases. Given the different techniques available for measuring market boundaries, and the different economic philosophies regarding government's role in regulating market activity (*see* CAPITALISM), debate over market definition methodology will likely continue.

See also *antitrust policy (US); geographic market; horizontal merger guidelines; merger guidelines, 1992–7*

Bibliography

Boyer, K. (1979). Industry boundaries. In T. Calvani and J. Siegfried (eds.), *Economic Analysis and Antitrust Law*. Boston: Little, Brown.

Breit, W. and Elzinga, K. (1996). *The Antitrust Casebook: Milestones in Economic Regulation*, 3rd edn. Fort Worth, TX: Dryden Press.

Church, J. and Ware, R. (2000). *Industrial Organization: A Strategic Approach*. Boston: McGraw-Hill Irwin.

Scherer, F. and Ross, D. (1990). *Industrial Market Structure and Economic Performance*, 3rd edn. Boston: Houghton-Mifflin.

Shepherd, W. and Shepherd, J. (2004). *The Economics of Industrial Organization*, 5th edn. Long Grove: Waveland Press.

Shepherd, G., Shepherd, H., and Shepherd, W. (2000). Sharper focus: Market shares in the merger guidelines. *Antitrust Bulletin*, **45**, 835 85.

US Department of Justice and Federal Trade Commission (April 2, 1992; revised April 8, 1997). *Horizontal Merger Guidelines*. Washington, DC: US Government Printing Office.

Waldman, D. and Jensen, E. (2001). *Industrial Organization: Theory and Practice*, 2nd edn. Boston: Addison-Wesley.

Werden, G. (1992). The history of antitrust market delineation. *Marquette Law Review*, **76**, 123 215.

market microstructure

Robert E. McAuliffe

This is an element of market transactions that economists tend to overlook: the actual processes and outcomes of exchange. How do specific trading mechanisms affect the distribution of prices? Market microstructure deals with this question by investigating how different trading mechanisms affect the prices observed when exchange occurs. Researchers investigate how price setting rules evolve in markets and basically how prices are determined in markets.

Economists have not focused on the determination of the EQUILIBRIUM price because they argue that the market determines the price. But the question remains: how do markets determine prices? Would the price change if the pricing mechanism were changed? Market microstructure examines these issues, normally in the context of asset markets for stocks and bonds.

In his work on intermediaries, Daniel Spulber (1999) found that intermediary firms in markets lower TRANSACTIONS COSTS. He finds that intermediaries can improve the gains from trade because they offer price commitments to both sellers and buyers of the good. Thus intermediaries can evolve and operate in a market even when direct exchange possibilities between buyer and seller exist. He also found that as transactions costs fell to zero, the intermediary price approached the market price but intermediaries set prices otherwise. When uncertainties in exchange are allowed or inventories, the importance of intermediaries is increased as they provide market making services to the economy.

Bibliography

Frankel, J. A., Galli, G., and Giovanni, A. (eds.) (1996). *Microstructure of Foreign Exchange Markets*. Chicago: University of Chicago Press.

Harris, L. (2002). *Trading and Exchanges: Market Microstructure for Practitioners*. Oxford: Oxford University Press.

O'Hara, M. (1997). *Market Microstructure Theory*. Oxford: Blackwell.

Spencer, P. D. (2000). *The Structure and Regulation of Financial Markets*. Oxford: Oxford University Press.

Spulber, D. F. (1999). *Market Microstructure: Intermediaries and the Theory of the Firm*. Cambridge: Cambridge University Press.

market performance

Gilbert Becker

Market performance measures the degree to which firms in a market achieve certain socially desirable goals, including EFFICIENCY and equity. Several types of efficiency have been identified. The first, **production efficiency**, occurs when suppliers minimize the AVERAGE TOTAL COST of producing a good, in both the SHORT RUN and the long run. By choosing the optimal combination of labor and other variable inputs (*see* ISOQUANT ISOCOST CURVES), managers can minimize the short run per unit cost of production. By achieving MINIMUM EFFICIENT SCALE, long run costs of production can also be minimized. As a result,

production efficiency enhances firms' profitabil ity. In addition, lower production costs and COMPETITION lead to lower prices, which benefit consumers. Moreover, this type of effi ciency benefits society on the whole by minim izing the amount of its scarce resources needed to produce the product.

A second type of efficiency is **dynamic effi ciency**, which focuses on technological pro gress. Here efficiency is achieved if market conditions induce a socially optimal amount of research and development, leading to a max imum level of innovation and invention. Schumpeter (1950) first proposed that large firm size and high industry concentration en hanced innovation. Many explanations have been advanced in favor of this hypothesis. These include the argument that ECONOMIES OF SCALE exist in research and development, and that there exists a disproportionate availabil ity of funds for large firms in imperfect CAP ITAL MARKETS. Also, the existence of economic profits in concentrated markets (*see* ECONOMIC PROFIT) has been argued as a factor enabling greater research. Other argu ments conflict with Schumpeter, maintaining that the incentive to innovate is diminished in markets where competition is imperfect. Cohen and Levin (1989) survey this literature and find that although there is a wealth of empirical stud ies of these hypotheses, the results are inconclu sive. They suggest that other variables affecting dynamic efficiency, such as the demand for in novation and the technological opportunity for innovation, need additional investigation.

A third type of efficiency is that of **allocative efficiency**. This is achieved when the amount of resources allocated to (and thus the number of goods produced in) each market maximizes society's welfare by reducing the DEADWEIGHT LOSS associated with MARKET POWER. This type of efficiency occurs in a market when every consumer who is willing and able to pay the MARGINAL COST of producing that good receives it. PERFECT COMPETITION is one MARKET STRUCTURE that reaches this result since the competitive market price is determined by the marginal cost of the last unit produced. MONOPOLY may also achieve this result if the firm is able to practice first degree PRICE DIS CRIMINATION.

The final performance criterion, **equity**, in volves, in part, a question of the fairness of the levels of prices and profits established in markets. The topic is quite controversial. By one definition, zero long run economic profit constitutes a fair result, since the buyers pay only for the economic costs of production, and the sellers have all of their costs covered, includ ing the owners' OPPORTUNITY COSTS. An alternative definition of fairness establishes that any level of economic profit is acceptable as long as there is a voluntary exchange between buyers and sellers in a market which is reasonably com petitive. This divergence of opinion as to an appropriate definition of fairness has resulted in differences in state and federal government policies concerning antitrust enforcement (*see* ANTITRUST POLICY (US)). As one example, Folsom (1990) demonstrates that during the 1980s the National Association of Attorneys General, fearing that market power would in crease the likelihood of unfair transfers of wealth, established more stringent standards for mergers and for vertical restraints of trade than those introduced by the federal govern ment.

MEASUREMENT

Most empirical studies investigating issues con cerning market performance have used one of two measures. Perhaps most prominent are studies using measures of firm and industry **economic profit**. If long run rates of return on assets or owner's equity are derived in part from the existence of market power, such profits may indicate a lack of market fairness. More over, these profits may be indicative of ineffi ciency resulting from this power if BARRIERS TO ENTRY or other factors foster an environ ment wherein maximum efficiency is not forced by competition.

Unfortunately, the use of profit rate data suffers from a number of measurement prob lems. Fisher and McGowan (1983) point out several of these flaws. For example, the correct profit levels will not be measured if the value of the firm's capital, or the DEPRECIATION thereof, is measured using accounting as oppo sed to economic definitions (*see* ECONOMIC DEPRECIATION). Similarly, the value of AD VERTISING and research and development

cannot be correctly associated with only the year in which the expenditure takes place. As a result, the measure of ACCOUNTING PROFIT reported for that year may diverge from that of the true economic profit. In addition, profit rates need to be adjusted for differences in risk across industries if a correct evaluation of the extent of excessive profit is to be made. Finally, profit rates suffer conceptually as a measure of performance since firms with high profits may be more efficient and thus demonstrate good performance.

A second tool used to measure performance is the **price cost margin**. One theoretically appealing example is the LERNER INDEX, which measures the difference between price and marginal cost as a percentage of price. While also indicating profitability, the extent of allocative inefficiency is directly measured in this index since the deadweight loss to society depends on the size of the gap between the marginal value that consumers give to the last unit consumed (price) and its marginal cost of production. Unfortunately, data on marginal cost are also difficult to obtain. As a result, AVERAGE VARIABLE COSTS are often used as a proxy. In addition, recent studies of market performance have attempted to estimate cost markups by examining pricing data in specific industries over time, or by introducing models of GAME THEORY. These methods suffer from the need to make numerous assumptions concerning demand and cost conditions, or about oligopolistic behavior. These and other issues continue to make the evaluation of market performance difficult.

See also *structure conduct performance paradigm; market power*

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Industrial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Cohen, W. and Levin, R. (1989). Empirical studies of innovation and market structure. In R. Schmalensee and R. D. Willig (eds.), *Handbook of Industrial Organization*. New York: North-Holland.

Fisher, F. M. and McGowan, J. J. (1983). On the misuse of accounting rates of return to infer monopoly profits. *American Economic Review*, **73**, 82 97.

Folsom, R. (1990). State antitrust remedies: Lessons from the laboratories. *Antitrust Bulletin*, **35**, 941 84.

Schumpeter, J. (1950). *Capitalism, Socialism, and Democracy*, 3rd edn. New York: Harper and Row.

market power

Gilbert Becker

Market power is generally defined as the ability of one or more firms to influence price. Greer (1992) broadens the definition somewhat to include "the ability to subdue rivals," including suppliers and potential entrants as well as customers. This power may arise from firms involved in either single or multiple markets. Market power stems from a variety of possible sources and has a number of potential consequences. It has been examined theoretically and estimated empirically by an ever growing number of authors.

The principal source of market power is size – that of the single firm alone, or a group of firms – relative to the overall size of the market. This size, in theory, generates power through its effect on market supply. For the individual firm in the case of MONOPOLY, the firm's size is equal to that of the market. As such the firm enjoys the ability to influence the market price since it supplies the entire market and can choose the point along the market demand curve where it wishes to operate. In contrast, in PERFECT COMPETITION the individual firm's size is so small that its output has no noticeable impact on the market supply and it becomes a *price taker*.

In antitrust cases concerning monopoly, once the market has been properly defined the next step traditionally has been to examine the market share of the alleged monopolist in order to estimate its market power. For example, Microsoft Corporation, in its recent antitrust trial, was found to control nearly 90 percent of its market, and thereby was seen as holding substantial market power (*see* MICROSOFT ANTITRUST CASE). Similarly, for groups of firms, as the collective market share of a small group rises, the potential for (and perhaps likelihood of) COLLUSION increases, again raising the ability to influence price (*see* OLIGOPOLY).

Current antitrust guidelines have broadened their focus beyond simple market share analysis since market share alone may not indicate sustainable market power. Recently, guidelines in the US and European Union (EU) have moved toward defining a market as the smallest area in which market power exists and a small but significant nontransitory increase in prices (SSNIP) can be sustained (*see* ANTITRUST POLICY (US); EU COMPETITION POLICY, 2004; MARKET DEFINITION).

The acquisition of market power can be achieved through any of several methods. Perhaps the most obvious approach is through the development of a product that is both unique in its characteristics and desired by consumers. Building a better mousetrap, or personal computer operating system for that matter, will ordinarily lead consumers to a firm's door, and increase its market share and thus its ability to influence price. A second means for expanding a firm's size and power is through a horizontal merger with a rival already within the market. Third, in order to sustain, if not create, market power, it is essential that BARRIERS TO ENTRY exist. These barriers, which are sometimes natural (i.e., extraordinary start up costs), sometimes government created (i.e., occupational licensing requirements), and in other instances are the result of deliberate STRATEGIC BEHAVIOR, all tend to retard the growth in the number of firms in a market, and thereby protect the above normal prices and profits of existing firms. Finally, while firm size relative to the market is the critical source of market power, a firm's absolute size may also play a role. Sheer size in terms of assets or number of employees may generate political power, which could in some instances help create or enhance market power. The evidence here is sporadic at best. Moreover, it is important for managers to be aware that for antitrust purposes, the courts have consistently held that large size alone is not onerous.

Several measures have been constructed that examine market power. One of the first theoretical models developed was the **Rothschild index**. This index is defined as the ratio of the slope of the individual firm's demand curve to that of the market demand. Following the analysis presented above, the ratio for a monopolist will have a value of 1 as the demand for its product is equal to the market demand. Overall the value of the index ranges between zero (for perfect competition) and 1, with higher values being associated with greater market power. The practical application of this index has been limited by the difficulty of estimating demand curves (*see* ESTIMATING DEMAND).

Three other measures have been more widely used in measuring market power. Each goes beyond the ability of (a) firm(s) to influence the market and instead focuses on the outcome of using that power which typically manifests itself in some measure of increased profit. The first of these three is the LERNER INDEX, measuring the extent of the divergence between price (P) and MARGINAL COST (MC). It is calculated as

$$L = \frac{(P - MC)}{P}$$

and, as was true of the Rothschild index, may range in value from zero for perfectly competitive firms to positive values with a limit approaching 1 in cases where price and marginal cost diverge. Once again, greater values for L indicate the presence of more market power. One of the difficulties in using this measure lies in the measurement of marginal cost, the data for which are not readily available. As a result, early studies used AVERAGE VARIABLE COSTS as an estimate for marginal cost. More recent studies estimate the value of MC econometrically (see Bresnahan, 1989; Church and Ware, 2000).

A third measure of market power, presented by Bain (1941), involves a direct measure of **excess profits**. Under perfect competition economic profits should equal zero in the long run (*see* ECONOMIC PROFIT), while the existence of market power in some imperfectly competitive markets should provide positive profits. Although ACCOUNTING PROFIT data are readily available and have been used in numerous studies following Bain's work, several difficulties arise in properly adjusting these data to reach a value for economic profits (*see* MARKET PERFORMANCE for details on measurement problems of this and other variables). In addition, positive economic profits may occur in the short run even in the absence of market power.

For example, exogenous shocks to market demand may temporarily increase price and result in short term profit even in the case of perfect competition. Consequently, in order to correctly use this measure of market power, these profits should not be examined with a single static estimate. Instead, a long term trend of sustained economic profits is necessary to establish the existence of market power.

A final measure of market power that has been used in some studies is **Tobin's q**. This measure, calculated as the ratio of the current market value of a firm's assets relative to the replacement cost of those assets, is an alternative method for measuring excess profits. Values of this ratio that exceed 1 indicate the existence of such profits. This measure has the advantage of not requiring an estimate of economic profits, as in Bain's measure, but does require the estimation of replacement cost, which in some instances may be quite difficult.

A great number of studies investigating the existence of market power have been undertaken over the past half century. In a survey of this literature Bresnahan (1989) identifies two classes of studies. The early studies followed Bain (1951) and the structuralist school. Tests of this school of thought commonly involved CROSS SECTION ANALYSIS, across several industries, of the effects of MARKET STRUCTURE on market performance. Here market power, arising from various structural factors, was expected to manifest itself in terms of firm and industry profitability, which was used as an indicator of performance. As noted above, the primary source of market power is firm size. As such, these tests examined the effects of single firm market shares and/or industry concentration (sources of power) on the value of the Lerner index or profit rates (the outcome of the use of power).

These studies found a consistently positive but weak link between size (market power) and profitability. Their conclusion that size caused profitability was hampered by the argument that the efficiency of large size may also be driving this profitability. Because of this, Bresnahan argues, a new class of studies identifiable as industry specific econometric case studies has arisen. In surveying the results of these studies he concludes that:

1 some concentrated industries possess significant amounts of market power; and
2 anticompetitive conduct is a significant cause of market power.

Numerous consequences may arise from the existence of market power. First, positive economic profits which result from the use of this power indicate a transfer of income from consumers to producers. Second, poor market performance, ranging from allocative inefficiency to a reduced rate of product innovation, are alleged to result from this power. Third, following Greer's definition, market structure can be changed if suppliers and rivals can be "subdued." The use of PRICE DISCRIMINATION, exclusive dealing, PREDATORY PRICING, and other forms of conduct often requires the existence of some power, but also may enable the further extension of that power across markets and over time.

See also *structure conduct performance paradigm*

Bibliography

Bain, J. S. (1941). The profit rate as a measure of monopoly power. *Quarterly Journal of Economics*, **55**, 271 93.

Bain, J. S. (1951). Relation of profit rate to industry concentration: American manufacturing, 1936 1940. *Quarterly Journal of Economics*, **65**, 293 324.

Bresnahan, T. (1989). Industries with market power. In R. Schmalensee and R. D. Willig (eds.), *Handbook of Industrial Organization*. New York: North-Holland.

Church, J. and Ware, R. (2000). *Industrial Organization: A Strategic Approach*. Boston: McGraw-Hill Irwin.

Greer, D. F. (1992). *Industrial Organization and Public Policy*, 3rd edn. New York: Macmillan.

Martin, S. (1994). *Industrial Economics*, 2nd edn. New York: Macmillan.

Maurice, S. and Thomas, C. (2002). *Managerial Economics*, 7th edn. Boston: McGraw-Hill Irwin.

market structure

Gilbert Becker

Market structure consists of those relatively fixed features of a firm's environment which identify the competitive nature of the industry. As such it is related to MARKET POWER and

MARKET PERFORMANCE and may be influ
enced by antitrust policy (*see* ANTITRUST
POLICY (US)). Its first element is the number
and size distribution of the sellers in the market.
As the number of sellers increases, the market
moves toward PERFECT COMPETITION and
market power tends to diminish. Several meas
ures of the size distribution of firms, including
various CONCENTRATION INDICES and the
HERFINDAHL HIRSCHMAN INDEX, empha
size the collective market shares of an industry's
leading firms. High values in these indices indi
cate a potential for market power. The market
share of a single firm may also help to explain the
market's structure and competitive conditions
(*see* PRICE LEADERSHIP). The second element
of market structure is the number and size dis
tribution of buyers, which is important as it
offers an indication of the extent of their coun
tervailing power. A proper count of the number
of sellers and buyers in any market requires
careful identification of the degree of product
substitutability between firms (*see* SUBSTI
TUTES) and the relevant GEOGRAPHIC
MARKET in which the competition takes place
(*see* MARKET DEFINITION).

A third vital element of market structure is the
condition of ENTRY. High BARRIERS TO
ENTRY are a central feature of MONOPOLY
and are common to OLIGOPOLY. The condition
of entry is important in understanding the com
petitive process in two ways. First, it helps to
explain the number and size distribution of firms
currently in the market. Second, it helps to
evaluate the potential for new competitors.

Some researchers cite this element of struc
ture as being uniquely important. Baumol, Pan
zar, and Willig (1982) maintain that in the
absence of any barriers to entry or exit (*see*
SUNK COSTS) markets become contestable and
the number of rivals, their size, and other struc
tural variables become irrelevant in determining
the outcome of market performance (*see* CON
TESTABLE MARKETS). Their research cites
examples such as the airline industry where
high resource mobility and low barriers to
entry into new geographic markets assure market
EFFICIENCY. Despite this, Shepherd (1984)
and others have argued that the number of
truly contestable markets is extremely limited,

and that the level of COMPETITION already in
existence in a market is of far greater importance
than the degree of potential competition.

Two other elements of market structure are
the degree of PRODUCT DIFFERENTIATION
and the extent of VERTICAL INTEGRATION
existing in the market. Both influence the nature
of industry costs and the STRATEGIC BEHAV
IOR among the rivals within the market. Both
may also play a role in determining the condition
of entry by increasing the costs and risk of entry.

The structure of a market is dependent on
several factors, the two most basic of which are
the underlying consumer demand and the indus
try's production cost conditions. Industry tech
nology which offers substantial ECONOMIES OF
SCALE relative to market demand may require
large firm size and greater market concentration
and thus limit the room for, and number of,
existing rivals. In addition, these circumstances
may limit the number of potential new entrants,
which may face cost disadvantages stemming
from an inability to produce at MINIMUM
EFFICIENT SCALE.

Market structure may also be influenced by
government policy ranging from patent laws and
licensing requirements, which influence entry,
to antitrust policy such as restrictions on
mergers (*see* HORIZONTAL MERGER GUIDE
LINES), which may influence the number and
size distribution of existing rivals. Finally,
LIMIT PRICING and other forms of strategic
behavior can be used by rival firms to alter the
structure of a market.

The work by Porter (1980) on competitive
strategy emphasizes the importance of market
structure to successful business management.
Five basic forces (including potential
entrants, substitute goods industries, and the
rivalry of materials suppliers) that exist in
every market are identified. Porter demonstrates
that from these forces evolve the competitive
strategies that firms must adopt in order to be
profitable. He argues that a sound evaluation
of market structure and market definition is
essential for managers in order to properly de
velop offensive and defensive strategies, assess
the company's strengths and weaknesses, and
examine its ability to cause changes in market
structure.

Analysis of market structure is also important in the development of industrial policy. Shep herd (1982) uses an analysis of market structure to investigate the extent to which goods in the US economy were produced in competitive markets. He uses structural elements to create categories such as "effective competition" (wherein industries have low concentration ratios, unstable market shares, and low entry barriers) and "tight oligopoly" (where concen tration ratios exceed 60 percent, and entry bar riers are medium or high). He finds that more than 75 percent of the economy's national income in 1980 was generated in markets that were competitive. Moreover, this percentage had increased sharply (from somewhat more than 50 percent) since the 1950s. Shepherd cites active government antitrust policy as the primary explanatory variable, and imports as a secondary variable causing this shift. The policy implications here, and surrounding market structure in general, are controversial (*see* STRUCTURE CONDUCT PERFORMANCE PARADIGM).

Bibliography

Adams, W. and Brock, J. (2001). *The Structure of Ameri can Industry*, 10th edn. Upper Saddle River, NJ: Pren tice-Hall.

Bain, J. S. (1956). *Barriers to New Competition*. Cam bridge, MA: Harvard University Press.

Baumol, W. J., Panzar, J. C., and Willig, R. D. (1982). *Contestable Markets and the Theory of Industry Structure*. San Diego: Harcourt Brace Jovano vich.

Carlton, D. W. and Perloff, J. M. (2000). *Modern Indus trial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Deutsch, L. (1998). *Industry Studies*, 2nd edn. Armonk, NY: M. E. Sharpe.

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.

Shepherd, W. (1982). Causes of increased competition in the U.S. economy, 1939 1980. *Review of Economics and Statistics*, **64**, 613 26.

Shepherd, W. G. (1984). "Contestability" vs. competi tion. *American Economic Review*, **74**, 572 87.

Shepherd, W. and Shepherd, J. (2004). *The Economics of Industrial Organization*, 5th edn. Long Grove: Wave land Press.

markup pricing

Kostas Axarloglou

This can be an optimal pricing policy where prices are set to cover all direct costs plus a percentage markup for profit contribution. It applies only for market structures where com panies have some MARKET POWER, and there fore are able to set the prices of their products (*see* MARKET STRUCTURE). In other words, it is used by *price setting* companies as opposed to *price taking* ones in a perfectly competitive market (*see* PERFECT COMPETITION).

Markup pricing is one of several different pricing policies companies use to price their products. For that, they calculate the cost per unit of output, to which they add a profit margin, or a price markup, to determine the final price of their product. The cost per unit is the variable production and marketing cost per unit of output plus the average overhead cost; the final price is then determined by adding a certain percentage margin to the unit cost.

The price markup is usually expressed as the difference between the final price and the mar ginal cost, as a percentage of either the price or the marginal cost. Here, the MARGINAL COST represents the per unit cost of production. The former method gives the markup on price, while the latter gives the markup on cost. If P repre sents the final price of the product, and MC its marginal cost, then the markup on price equals to $\dfrac{P - \text{MC}}{P}$, and the markup on cost is $\dfrac{P - \text{MC}}{\text{MC}}$.

Although the markup pricing policy is rela tively easy to apply, there are some problems associated with its use. First, the overhead cost per unit is calculated based on the expected overhead cost divided by the "normal" amount of the product the company expects to produce. However, in many cases, the actual production deviates from the expected volume of produc tion, and this becomes a source for cost miscal culations. Also, in calculating the per unit cost of production, companies usually use actual ex penses only, and they ignore the OPPORTUN ITY COSTS of their resources. Consequently, they may underestimate their cost of production. Finally, since companies do not change their prices very frequently, they need to determine

the optimal price that will cover the period of time until the next price adjustment.

Economic theory suggests that the pricing rule, which allows companies to maximize their profits (*see* PROFIT MAXIMIZATION), is $P = \dfrac{1}{1 + \frac{1}{e_p}}$ MC, where e_p is the price ELASTICITY of the firm's demand for its product (when there are no strategic interactions between firms; *see* GAME THEORY and OLI GOPOLY for a discussion of interaction between rival firms). Obviously, the profit maximizing price depends on the marginal cost, and it is inversely related to e_p, the price elasticity of demand (notice that e_p is always a negative value). In other words, companies with very elastic market demand (large absolute value of e_p), and therefore a market demand very sensitive to changes in prices, will charge a price close to their marginal cost. Their price markup will be low, since high prices might result in a significant decline in their sales. In a perfectly competitive market, where e_p is a very large number (in extreme cases it approaches ∞), P is very close to MC, or the price markup is close to zero. From the optimal pricing equation above, the price to cost markup can be derived as $\dfrac{P - \text{MC}}{P} = -\dfrac{1}{e_p}$, which is called the LERNER INDEX, introduced by Abba Lerner in 1934. Obviously, the price markup is inversely related to the price elasticity of market demand. High e_p makes the company very aggressive in pricing its products and willing to accept a low price markup as a way to stay competitive and preserve its market share. On the other hand, low e_p (demand is relatively price inelastic) implies large profit maximizing price markups. In such a case, the company has some market power, and for that reason it can charge a rather large price markup over its marginal cost. (For the factors that influence e_p, see the discussion in the entry on elasticity).

The above discussion suggests that the price markup is closely related to the degree of market power a company can exercise. More competitive market structures are usually characterized by lower price markups. Domowitz, Hubbard, and Petersen (1986), by using data from US industries between 1958 and 1981, found that those industries with high concentration (with a four firm concentration ratio in the range of 80 to 100) show a markup on cost ratio of 0.32, while those industries with low concentration (with four firm concentration ratio in the range of 0 to 20) show a lower markup on cost ratio of about 0.23, as expected. Also, Hall (1986), using data from 48 US industries in manufacturing for both durable and non durable products, shows that in most of these industries price markups are significantly different from zero, or that in US manufacturing overall, companies seem to have significant market power.

Rotemberg and Saloner (1986) and Rotemberg and Woodford (1991, 1992) present evidence that price markups are correlated with the changes of economic activity during the different phases of the business cycle. They usually increase during economic recessions and decline during economic booms. Finally, Axarloglou (1994) finds some evidence of the same behavior in the publishing industry.

Bibliography

Axarloglou, K. (1994). Product variety and countercyclical price markups: An empirical assessment. Working paper, Fuqua School of Business, Duke University.

Domowitz, I., Hubbard, R. G., and Petersen, B. C. (1986). The intertemporal stability of the concentration margins relationship. *Journal of Industrial Economics*, **17**, 1 17.

Hall, R. (1986). Market structure and macroeconomic fluctuations. *Brookings Papers on Economic Activity*, **2**, 285 322.

Greer, D. F. (1992). *Industrial Organization and Public Policy*, 3rd edn. New York: Macmillan.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Lerner, A. (1934). The concept of monopoly and the measurement of monopoly power. *Review of Economic Studies*, **1**, 157 75.

Rotemberg, J. and Saloner, G. (1986). A supergametheoretic model of price wars during booms. *American Economic Review*, **76**, 390 407.

Rotemberg, J. and Woodford, M. (1991). Markups and the business cycle. *NBER Macroeconomic Annual*, 63 129.

Rotemberg, J. and Woodford, M. (1992). Oligopolistic pricing and the effects of aggregate demand on economic activity. *Journal of Political Economy*, **100**, 115 207.

maximin criterion

Eduardo Ley

Suppose that, when a manager is faced with a decision, the payoffs from choosing any of the different strategies available will depend on circumstances beyond her control (e.g., stochastic events, other people's strategies, etc.), which we shall refer to as "states of the world." If we rank the strategies available by each strategy's payoff in the worst case scenario, then the highest ranked is a maximin strategy. A maximin strategy has the property that the worst that can happen with it is no worse than the worst that could happen with any of the other strategies available. In other words, this is a conservative criterion for choosing a strategy which is the best in the worst possible scenario. As an example, suppose that there are two available decisions, d_j, and three possible states of the world, ω_i, with the probabilities and payoffs shown in the following table:

	d_1	d_2	$\mathrm{Prob}(\omega_i)$
ω_1	25	1	0.50
ω_2	10	15	0.35
ω_3	−5	5	0.15

The worst scenario for decision 1 has payoff −5, while the worst payoff with decision 2 is 1. Since $1 > -5$, decision 2 would be chosen by the maximin criterion despite the fact that decision 1 has a higher expected payoff (*see* EXPECTED VALUE). Note that the maximin criterion does not take into account the probabilities of the different states of the world. (In public economics the maximin criterion refers to the welfare criterion, which mandates maximizing the utility of the individual who is worst off.)

Bibliography

Raiffa, H. (1968). *Decision Analysis: Introductory Lectures on Choices under Uncertainty*. Reading, MA: Addison-Wesley.

merger guidelines, 1992 7

Gilbert Becker

These guidelines are the third set of guidelines offered to the US business community in the past quarter century. They are an effort to clarify the circumstances under which the current administration will challenge a merger which it believes to be in violation of the CLAYTON ACT. The guidelines offer a five step approach to the analysis of a merger:

1 the definition of the market and measure of industry concentration (*see* CONCENTRATION INDICES);
2 the evaluation of the potential adverse effects of the proposed merger;
3 the analysis of the condition of ENTRY into the market (*see* BARRIERS TO ENTRY);
4 the existence of potential efficiencies from the merger (*see* EFFICIENCY); and
5 the examination of the circumstances in the case where one of the merging firms is failing.

The geographic and product market boundaries are defined using the method known as the 5 percent rule, or SSNIP – a small but significant nontransitory increase in prices (*see* GEOGRAPHIC MARKET; MARKET DEFINITION). The market is then classified as being highly concentrated (if the post merger HERFINDAHL HIRSCHMAN INDEX, HHI, has a value greater than 1,800), moderately concentrated (if, after the merger, $1,000 < \mathrm{HHI} < 1,800$), or unconcentrated (if the post merger HHI < 1,000). While a challenge in the latter case is unlikely, in general the likelihood of a challenge increases, all else constant, for higher levels of concentration and for larger changes in the level of concentration. The numerical ranges used in the 1982 guidelines remain in force here, but new detail is presented (i.e., concerning market definitions where PRICE DISCRIMINATION exists), and a broader description of economic evidence being used by the antitrust agencies is offered.

The second step involves numerous tests which evaluate factors inherent in the industry that facilitate or limit the potential for anticompetitive effects to arise as a result of the proposed merger. Here the potential abuse of MARKET POWER is seen as having two possible sources:

1 coordinated efforts among the sellers in the market;
2 unilateral actions by the newly merged, and now larger, firm.

Evidence indicating a diminished capacity for the abuse of market power will tend to lessen the likelihood of a challenge. When the reverse is true the likelihood of a challenge is enhanced. For example, product standardization in an industry and free flowing information about rivals' competitive actions are two factors cited as being conducive to COLLUSION. Similarly, the closeness of product attributes of the merging firms' products is cited as facilitating a successful unilateral post merger increase in prices in some circumstances (*see* PRODUCT ATTRIBUTES MODEL).

The 1992–7 guidelines recognize that any adverse effects of a merger may be offset by the entry of new rivals. Thus the third step of the new guidelines examines the condition of entry into the market and now recognizes the particular importance of three factors:

1 the timeliness of the new entry;
2 the likelihood that it will occur; and
3 the sufficiency of that entry in controlling abuses of market power.

Fourth, the extent to which a proposed merger may offer a method for achieving ECONOMIES OF SCALE or other cost savings is considered as a possible defense in cases which might otherwise be challenged. The antitrust agencies will examine potential efficiencies stemming from the merger, and the new guidelines outline several possible sources of cost savings that may be considered. The 1997 revision of the guidelines is focused on the section concerning efficiencies and emphasizes that only *merger specific efficiencies* will be considered. These are efficiencies which are likely to be achieved by the merger and unlikely to be achieved in its absence. If these efficiencies are *cognizable* – verifiable and not arising from anticompetitive reductions in output – then they will be considered. Moreover, the greater the potential adverse effect of the merger, i.e., as measured by increased concentration, the greater must be the efficiencies if the merger is to be deemed acceptable.

Finally, the guidelines recognize that the threat of increased market power is minimized in cases where one of the merging firms is in imminent danger of failure. Here the guidelines consider four situations wherein a challenge to the proposed merger is unlikely, including:

1 where the likelihood of the failing firm to successfully reorganize under current BANKRUPTCY laws is poor;
2 where there is an absence of alternative merging partners with whom there exists a decreased potential danger to COMPETITION arising from a merger.

These guidelines are largely an extension of those presented in 1982–4. It is generally accepted that these guidelines have brought more careful economic analysis into the process (see Scheffman, 1993, for a symposium of views). Both sets of guidelines are also widely accepted as being more lenient toward firms that wish to merge than those introduced in 1968. In an examination of recent antitrust activity, Lande (1994) finds that of 61 mergers challenged by the Federal Trade Commission between 1987 and 1992, only one involved a merger with a post merger HHI value less than 2,000. In addition, only four challenges involved mergers where the change in the HHI was less than 400 points. He cites this and other evidence as indicative of the attenuation of aggressiveness in antitrust enforcement.

Heightened concerns by business over increasing foreign competition during the past decade may have fueled the discussion of a need for greater leniency in the guidelines. While the two most recent sets of guidelines specifically recognize the need to include foreign imports in the definition of the market, they also clearly recognize that trade restrictions may limit the ability of foreign firms to respond to domestic price increases.

See also *antitrust policy (US); EU merger guidelines, 2004; EU merger policy, 2004; geographic market; horizontal merger guide lines*

Bibliography

Lande, R. (1994). Beyond Chicago: Will activist antitrust arise again? *Antitrust Bulletin*, **39**, 1 25.

Ordover, J. and Willig, R. (1993). Economics and the 1992 merger guidelines: A brief survey. *Review of Industrial Organization*, **8**, 139 50.

Scheffman, D. (1993). Symposium on new 1992 merger guidelines. *Antitrust Bulletin*, **38** (3).

Shepherd, G., Shepherd, H., and Shepherd, W. (2000). Sharper focus: Market shares in the merger guidelines. *Antitrust Bulletin*, **45**, 835 85.

US Department of Justice and Federal Trade Commission (April 2, 1992; revised April 8, 1997). *Horizontal Merger Guidelines*. Washington, DC: US Government Printing Office.

US Federal Trade Commission Bureau of Competition (1992). *How 1992 Guidelines Differ from Prior Agency Standards*. Washington, DC: US Federal Trade Commission.

Microsoft antitrust case

Gilbert Becker

One of the champions of the computer software revolution over the past two decades, the Micro soft Corporation has also been the center of controversy regarding many of its business strat egies. Since the mid 1990s the company has been the subject of a number of antitrust actions, brought by the federal and state governments in the US, by groups of consumers, by rival firms, and by the European Union (EU). At issue are a number of alleged anticompetitive business practices, the injury which these practices may have brought on the market and its various parties, and appropriate public policy toward these acts. Given the magnitude of the impact the new technology has had on numerous markets and the complexity of the issues in question, these cases are seen as an important opportunity to reexamine major antitrust stat utes such as the SHERMAN ACT, perhaps in light of new economic thinking about new forms of COMPETITION evolving in the twenty first century.

BACKGROUND AND CURRENT STATUS OF THE CASE

Microsoft (MS) is perhaps best known for its development and sale of the product called Windows, which is the operating system cur rently running on approximately 90 percent of all personal computers in the US. The company has also developed numerous other computer software applications programs (e.g., Microsoft Word for word processing, Excel for spreadsheet analysis, etc.) which run on the Windows system, some of which are at the center of the debate. Other programs, such as Internet Ex plorer for accessing the Internet, have been writ ten into or have become part of upgraded versions of the Windows operating system.

In 1995, as the result of an investigation by the US Department of Justice (DOJ), Microsoft entered into a consent decree in which it agreed not to tie the sale of other products to that of Windows. In 1997, the DOJ filed suit against Microsoft, claiming that the company had vio lated the decree by BUNDLING its web browser, Internet Explorer (IE), with Windows. In June of 1998, shortly before a US Court of Appeals ruled in favor of Microsoft, the DOJ in conjunc tion with 20 US states filed a broadened antitrust suit against the company, claiming illegal mon opolistic practices.

That case was initially decided in US District Court in April of 2000 (see *US* v. *Microsoft Corp.*, 1999, 2000). Judge Jackson's three central findings were that Microsoft had violated Sections 1 and 2 of the Sherman Act by:

1 illegally monopolizing the operating system market;
2 illegally attempting to monopolize the web browser market; and
3 illegally tying Windows and IE.

Subsequently, a US Circuit Court of Appeals overturned the last two findings of Judge Jackson (see *US* v. *Microsoft Corp.*, 2001).

MARKET DEFINITION AND MARKET POWER

Numerous economic and legal issues have arisen over which the two parties (Microsoft and the government) and the two courts (District and Circuit) have often disagreed. The first issue to

arise is the proper definition of the market (*see* MARKET DEFINITION). This issue is critical in that illegal monopolization and illegal tying req uire the demonstration of the existence of MARKET POWER, and the abuse of that power in the relevant market(s). Thus, proper market boundaries must be established before any fur ther economic analysis can take place.

Franklin Fisher, the principal economic wit ness for the DOJ during the District Court trial, used the new theoretical approach to market definition provided in the government's HORI ZONTAL MERGER GUIDELINES. (Much of what follows is based on the testimony of Fisher; see Fisher and Rubinfeld, 2001, for a detailed economic analysis.) He argued that goods which serve to constrain a hypothetical monopolist's market power (*see* SUBSTITUTES) should be included in a proper definition of the market. The DOJ held that two separate economic markets exist: (1) Intel based PC operating systems and (2) web browsers. In the *Intel based PC operating system* (OS) market, Windows held approximately 90 percent market share, well above the Supreme Court's previously es tablished standard for demonstrating the exist ence of MONOPOLY. Here, hand held substitutes (e.g., Palm Pilots) were seen as out side the market, due to their low CROSS ELAS TICITIES, and non Intel PCs (e.g., Apple) were considered to be a limited damper on Micro soft's power. Moreover, some original equip ment manufacturers (OEMs) of PCs testified to the lack of alternatives to purchasing Windows for their machines. Fisher also argued that browsers were a complementary good (*see* COM PLEMENTS) to operating systems, since they initially held a separate price and market of their own.

As for Microsoft's market power, the govern ment's case relies heavily on the theory of NET WORK EXTERNALITIES and the existence of an applications programming barrier to entry (*see* BARRIERS TO ENTRY). Independent program mers have written tens of thousands of software programs that will run on the Windows operat ing system by relying on Microsoft written ap plications programming interfaces. These are sets of commands by which applications connect with the personal computer. Fisher argues that the operating system market is a "winner

take most" market, since consumers will gravi tate to that system with the most programs, given that the more consumers using a program (e.g., Word, Instant Messenger), the more valu able it becomes to other potential users (to "join the network"). In addition, programmers will gravitate to writing almost entirely for that oper ating system with the most customers, since that is likely to be their most profitable strategy. Reinforcing this focus by programmers are two additional factors:

1 most programming costs are FIXED COSTS leading to dramatic programming ECON OMIES OF SCALE; and
2 costs for rewriting the program to run on another system are often high.

Thus, unless a new operating system can attract a huge volume of applications, a network effects applications barrier yields market power for the dominant operating system.

Microsoft's position, as presented by its prin cipal economics witness Richard Schmalensee, was that the proper market to be examined was that of *software platforms*, which are sets of code upon which applications can be used. (See Evans, Nichols, and Schmalensee, 2001, for a detailed description of Microsoft's economic de fense in the case.) While platforms also result in winner take all network effects, Schmalensee argued that Microsoft's Windows market pos ition was tenuous, since platforms can be de veloped from other sources, such as Sun Microsystem's Java or by using a browser (e.g., Netscape's Navigator) to create a web based platform. Both of these so called middleware applications threatened the Windows monopoly, albeit in different ways. With respect to Nets cape, programmers could write software that would run on its Navigator, which runs on other operating systems and threatened to become a platform substitute replacing Windows. Java, on the other hand, was an open architecture programming language that was designed to run on any operating system. This meant that (potentially many) new operat ing systems, each running programs written in the Java language, could be developed as substi tutes for Windows and thereby threaten to "commoditize" the once essential product.

Moreover, one "killer" application for any platform could lead to widespread demand by consumers, and a flood of new applications being written for that platform. Thus, while Microsoft was not competing head to head with these firms in the Intel based operating system market, Schmalensee argued that Microsoft's market power is more limited when measured in a framework of dynamic competition in the platform market. In this setting, innovation by "new and unforeseen" competitors poses a con tinuing threat to the existence of even a firm with a dominant market share. Using the DOJ's theory of a barrier protected monopolist, he demonstrated Microsoft's limited power by esti mating Microsoft's correct profit maximizing monopoly price to be hundreds of dollars greater than the price that was actually charged for Windows (*see* PROFIT MAXIMIZATION). Finally, Microsoft's position with respect to its IE was that the browser is an important innov ation which is a natural addition to an operating system since it allows consumers access to web based information and can be used by Windows to provide upgrades, and therefore is not a sep arate product.

In addition, Schmalensee argued that, for sev eral reasons, there does not exist an applications barrier to entry in the market. First, widespread research and development and investment in intellectual property exist across the industry. Second, thousands of software writers exist in the market and barriers to entry in software programming are low. Third, many writers do offer programs for less used platforms, espe cially where buyers (e.g., businesses) may be willing to pay more. Finally, since applications are a desirable consumer product, identifying the existence of many applications as a barrier to entry is comparable to identifying improve ment of product quality as a barrier.

Anticompetitive Behavior

The government argued that Microsoft was in volved in numerous predatory and exclusionary acts designed to maintain its operating system monopoly by protecting its applications barrier to entry. First, Fisher argued that Microsoft tied its browser to the sale of Windows in order to monopolize the browser market. This was not done to ultimately charge a monopoly price for

IE. Instead Microsoft's intent, according to Fisher, was to protect its operating system mon opoly, since it feared that Netscape's Navigator, a complementary good to the operating system, could develop into a platform for users and pro grammers which could ultimately run on any operating system.

In addition, the government alleged that Microsoft attempted to foreclose competition in the browser market in several ways, including:

1   by attempting to illegally divide the browser market by soliciting Netscape to not produce a browser that would run on Windows 95;
2   by annual expenditures of $100 million for development of IE, combined with giving the browser away for free, demonstrating obvious predatory behavior;
3   by restricting OEMs from removing IE or its desktop icon, which, combined with PC manufacturers' desire to load only one browser, excluded Netscape Navigator from most new shipments;
4   by using its Windows' market power to induce Internet Service Providers (ISPs) to limit their marketing and distribution of non Microsoft browsers; and
5   through agreements with other software de velopers desiring Windows support (e.g., Intuit) to exclusively deal with Microsoft.

Fisher argued that each action, taken on its own merit, was contrary to Microsoft's short term profit and could thus only be explained as antic ompetitive efforts to sustain its monopoly.

The government also identified efforts by which Microsoft attempted to directly protect its operating system monopoly. These included:

1   making threats against Intel Corp. and Apple Computer Co. to withhold Microsoft's sup port of their products. These threats forced each firm to agree to end work on new technology (e.g., native signal processing, audiovisual streaming) which potentially threatened the Windows monopoly, and to limit support of Netscape Navigator and Sun Microsystem's Java.
2   Microsoft licensed Java from Sun, then created a "polluted Java" of their own which would cause some independent

software developers' programs to run prop erly on Windows, but not on non Windows operating systems. This threatened to elim inate a main selling point of Java, which was that it was designed to run on any operating system.

The government argued that these acts harmed markets in several ways. First, consumers were harmed by the suppression of new technology and innovations not reaching the market. More over, even if Microsoft later chose to incorpor ate a version of the new technology in upgrades of Windows, the decision as to which innov ations should be allowed access into the market should not be determined solely by Microsoft's interests. Second, consumers and PC manufac turers were unable to choose which browser would be preinstalled, nor were they able to purchase a browserless PC. Third, consumers faced significant time costs and other TRANS ACTIONS COSTS if they wished to download Netscape Navigator. Fourth, browser market rivals of Microsoft were being foreclosed from a fair market test of their products based on the quality and features offered to consumers. Finally, and perhaps most importantly, Judge Jackson, in his Findings of Fact, cited harm from Microsoft's signal that it "will use its market power and immense profits to harm any firm that pursues initiatives that could in tensify competition against one of Microsoft's core products" (*US* v. *Microsoft Corp.*, 1999: 412). In sum, the DOJ's position was that what ever benefits were brought to market by Micro soft's development of IE, they did not exonerate Microsoft from its illegal acts designed to maintain its monopoly. Antitrust action is appropriate as it is designed to protect consumers by *protecting competition*.

Microsoft's position with respect to charges of predation and exclusion hinges on the fact that the platform market is a winner take all environ ment. In this setting Schmalensee argues that each rival may set its price below variable cost, violating the standard test for predation (*see* PREDATORY PRICING), expecting to win even though only one firm will succeed. He further argues that while IE's pricing brought no rev enue from IE itself, the strategy was essential for the survival of the Windows platform and of the firm. Microsoft's expenditures of hundreds of millions of dollars on IE, and its subsequent zero pricing for its product, along with its other ex clusionary strategies, were acts of competition. As such these acts should not be considered illegal. The inevitable victory of only one firm means that the traditional measure for predation is inappropriate, in large part since it condemns the winner, or as Schmalensee stated: "success, monopolization and exclusion are one and the same" (Evans et al., 2001: 212). As an alternative he offers that a new antitrust standard for preda tion in winner take all markets be the direct examination of the benefits versus costs to con sumers of the business actions taken. Here Schmalensee argues that Microsoft's consider able expenditure on IE clearly resulted in a de sirable product benefiting consumers, as did the innovation race itself.

In addition, Microsoft argued that many of its actions did not significantly impact the market in question. Its CONTRACTS did not preclude PC manufacturers from installing more than one browser, and Netscape and other rivals offered free downloads over the worldwide web. More over, Netscape was able to reach over 100 million potential customers with a CD mailing which only marginally raised its overall costs. Also, other business pressures applied by Microsoft to members of the industry (e.g., ISPs, Apple) did not succeed. In sum, Microsoft held that competition in the market continued to be healthy. Moreover, it maintained that the anti trust case was inappropriate in that it would stifle future innovation by Microsoft and only served to *protect competitors* who were unsuccess ful in the market.

The Courts concluded that Microsoft did indeed violate the antitrust laws. The final rem edies in this case have yet to be determined and remedies in other antitrust lawsuits against Microsoft are pending. Their implementation may prove to be a force in shaping the industry, and also future public policy. Any legislative changes (or judicial reinterpretation of the cur rent laws) which arise will undoubtedly continue to be debated.

See also *antitrust policy (US); capitalism; EU competition policy, 2004; Microsoft antitrust case: remedies*

Bibliography

Evans, D., Nichols, A., and Schmalensee, R. (2001). An analysis of the government's case in *US* v. *Microsoft*. In W. Commanor (ed.), Symposium: Economists examine Microsoft its concepts, findings, and remedies. *Anti trust Bulletin*, **46**, 163 252.

Fisher, F. and Rubinfeld, D. (2001). *US* v. *Microsoft*: An economic analysis. In W. Commanor (ed.), Symposium: Economists examine Microsoft its concepts, findings, and remedies. *Antitrust Bulletin*, **46**, 1 72.

US Circuit Court, District of Columbia (2001). *US* v. *Microsoft Corp.*, 253 F.3d 34 (D.C.Cir. 2001). Washington, DC: US Government Printing Office.

US District Court, District of Columbia (1999). *US* v. *Microsoft Corp.*, 98 CIV.1232 Court's Findings of Fact. Washington, DC: US Government Printing Office.

US District Court, District of Columbia (2000). *US* v. *Microsoft Corp.*, 98 CIV.1232 Court's Conclusions of Law. Washington, DC: US Government Printing Office.

Waldman, D. and Jensen, E. (2001). *Industrial Organiza tion: Theory and Practice*, 2nd edn. Boston: Addison-Wesley.

Microsoft antitrust case: remedies

Gilbert Becker

In the monopoly case against Microsoft Corpor ation (*see* MICROSOFT ANTITRUST CASE), Judge Jackson's decision concerning remedies for the illegal MONOPOLY called for both struc tural and behavioral relief (see *US* v. *Microsoft Corp.*, 2000). It first called for splitting Micro soft into two firms, with one firm responsible for the Windows operating system and the other owning the Microsoft applications software (i.e., Microsoft Office, Internet Explorer, etc.). The rationale behind this split was the expect ation that the first firm would continue to innov ate in the operating system market, while the second firm would have incentive to make the applications functional for other operating systems, thus allowing new competitors to enter and compete in the operating system market where Microsoft currently holds a mon opoly position. His decision also restricted Microsoft's conduct in several ways, including placing restrictions on Microsoft from making exclusive deals with Internet Service Providers

(ISPs), and ending its anticompetitive restric tions on PC manufacturers.

On appeal, the US Circuit Court concurred with the US Department of Justice's (DOJ) MARKET DEFINITION and affirmed the lower court's findings of illegal exclusionary and predatory monopolistic practices (see *US* v. *Microsoft Corp.*, 2001). It overturned the pro posed structural remedy, in part due to the se verity of the remedy, the uncertainty of its success, and the brevity of the hearings held on the matter. It remanded the case to the lower court for more complete hearings into an appro priate remedy.

Past antitrust remedies have been designed to end anticompetitive activity, promote a more competitive environment, and punish antitrust offenders. The existence of NETWORK EXTER NALITIES in industries such as computer soft ware may call into question the validity of applying traditional remedies because of possible conflicts arising in achieving these goals. Debate surrounding the question of structural relief in the Microsoft case centers in large part on the economics of these network effects in the industry.

Critics of the government's position on break ing up Microsoft argue that these effects make the District Court's (and other) structural rem edies harmful to consumers, software develop ers, and the economy as a whole. The core of the argument is that increasing the number of com petitors in the operating system market will result in a loss of some network benefits to con sumers. This happens in part because new soft ware applications ordinarily will not run on all operating systems, and as a result individuals will not be as able to share work or otherwise communicate with one another. Moreover, com petition among firms supplying operating systems may take the form of PRODUCT DIF FERENTIATION and lead to increasing incom patibilities across systems. Liebowitz (1999) estimates that software producers would face $30 billion in additional costs for development and marketing if they chose to adapt their existing software to conform to as few as two new operating systems. He argues that these costs would lead to higher software prices for consumers, or a reduced number of software firms as some may not be able to survive the

competitive problems of higher costs and a more fragmented set of consumers.

Microsoft proponents also argue that the firm's monopolizing conduct itself should not be subject to antitrust punishment because the natural result of rivalry in a market having net work effects is a winner take all outcome, i.e., monopoly. The traditional standards for SHER MAN ACT violations, which make exclusionary and predatory behavior illegal, are seen as pro tecting only unsuccessful competitors, and not protecting COMPETITION or consumers. Thus such standards are inappropriate in markets having significant network effects (see Rule, 1999).

A central question surrounding the validity of these arguments is whether achieving network effects inevitably requires dominance by a single firm. Proponents of structural relief argue that there is no compelling theoretical proof of this. Commanor (2001) points out that after the breakup of AT&T in the 1980s telephone service compatibility was maintained among the rivals in this market, which also yielded substantial net work effects. In addition he argues that while some economic models of OLIGOPOLY can sup port the claim that product differentiation may create incompatibilities in a restructured market, other models predict convergence toward prod uct standardization. Levinson, Romaine, and Salop (2001) expand this argument, showing that firms producing rival operating systems would each have strong incentives to maintain compatibility across their basic systems, to maxi mize network benefits for their customers, and thereby maximize their own sales and profits. While this core compatibility is being main tained, each firm simultaneously has incentives to add new features to its own operating system. Moreover, exclusionary and predatory behavior harms consumers and the competitive process if it unnecessarily results in one firm deciding which operating system innovations will be offered as choices to consumers.

In November 2001, the DOJ and nine states reached an out of court agreement as to remedies (see *US* v. *Microsoft Corp.*, 2001). These remed ies abandoned structural relief and focused en tirely on behavioral solutions. The District Court, as required by law, examined the pro posed settlement and simultaneously held hearings concerning an alternative proposal offered by nine states which had refused to sign the agreement drafted by the DOJ. Debate between the DOJ and these states concerned the types and strength of conduct remedies to be imposed. The DOJ, now under the Bush administration, had required what some considered to be only mild concessions from Microsoft, including:

1   allowing greater freedom for PC manufac turers to install non Microsoft software and remove Microsoft desktop icons, and pro hibiting retaliation against those who do;
2   prohibiting Microsoft from entering CON TRACTS which require exclusive support of its products;
3   requiring increased disclosure of technical data aiding programmers to design programs that will fully function on Windows.

The remaining states voiced concerns that the new administration's settlement was too lenient, difficult to enforce, and carried too many ex emptions for it to effectively constrain Micro soft's monopoly power. Moreover, they cited examples of Microsoft continuing some of its recently proscribed business practices in prod uct lines that have emerged since the conclusion of the initial trial. These states requested numer ous additional remedies, including:

1   requiring Microsoft to sell an unbundled ver sion of Windows (*see* BUNDLING), leaving a more openly competitive environment for rivals selling various software applications such as browsers, media software, and the like;
2   requiring Microsoft to license the source code of Internet Explorer (IE) in order to promote its use by rivals as a platform for their products;
3   forcing Microsoft to write its Office software to run on competing operating systems such as Linux, while restoring its support of the Java programming language; and
4   appointing a special master to monitor Microsoft's business behavior.

Finally, they asked for the courts' findings and remedies to apply to new and future markets in which Microsoft is involved.

Microsoft remained opposed to each of these changes, arguing that the unbundling of Windows would be difficult at best, inefficient, and harmful to consumers. Moreover, it would restrict the company's right to innovate. It fur ther argued that the states' proposals constituted a confiscation of intellectual property and would force Microsoft out of the market.

In November of 2002, the District Court accepted the DOJ settlement proposal as being in the public interest, and largely rejected the remaining states' case. Currently, one state, Massachusetts, remains actively opposed to the settlement agreement, and has filed an appeal with the Circuit Court. The full Circuit Court of Appeals was due to meet in late 2003 to hear this appeal and a separate appeal filed by a con sortium of Microsoft's rivals in the markets for server software and other new products. This group has convinced the court to examine the question of whether the District Court erred by not considering alleged antitrust violations by Microsoft in these new markets when handing down its decision regarding the settlement agreement concerning earlier antitrust viola tions.

Antitrust law also allows for monetary dam ages to be collected by individuals or firms who initiate private lawsuits. The court's conclusion in US v. Microsoft Corp. (2001), that Microsoft was guilty of illegal activity, dramatically simpli fies plaintiffs' cases for damages. Injured parties need only to demonstrate that the injury oc curred as a result of Microsoft's actions and convince the court of a specific dollar amount of damages. The court may then impose treble damage awards.

Consequently, Microsoft faces numerous damage suits. One important case is a set of over 100 class action lawsuits brought on behalf of consumers who allegedly paid excessive amounts for Windows. An initial out of court settlement of this case, whereby Microsoft offered several hundred million dollars in cash, free software, computers, and training to public schools, was rejected by the courts as potentially improving Microsoft's power in the education market (see Buckman and Kulish, 2002). In two additional cases, America on Line (AOL), which now owns Netscape, and Sun Microsystems, which owns Java, both of whom were cited by the Circuit Court as having been injured, filed lawsuits to collect for damages. In July of 2003, AOL and Microsoft reached a settlement in which Microsoft would pay AOL $750 million in damages.

Finally, the European Union (EU) is also investigating Microsoft for violations of its anti trust laws concerning "abuse of a dominant pos ition" (see EU COMPETITION POLICY, 2004). It is also examining Microsoft's efforts to use its Windows monopoly to dominate the market for corporate server software, and its bundling of Windows with its media software. The European Commission, which in recent years is widely considered to have taken a more aggressive antitrust stance than its US counter parts, is considering requiring that Microsoft provide an unbundled version of Windows. It is armed with considerable investigative powers, and also holds, as perhaps its most powerful remedial weapon, the ability to impose a penalty of up to 10 percent of a firm's annual worldwide revenues for violations of the law. In addition it maintains other antitrust remedies similar to those used in the US such as cease and desist orders. The final outcomes of the EU investi gation and the various cases in the US concern ing Microsoft may have a significant impact on antitrust policy in the US and worldwide in the foreseeable future.

See also *antitrust policy (US); antitrust remedies (US); Microsoft antitrust case; Sherman Act*

Bibliography

Buckman, R. and Kulish, N. (2002). Judge tosses Micro soft consumer class-action deal. *Wall Street Journal*, **244**, A3.

Commanor, W. (2001). The problem of remedy in mon opolization cases: The Microsoft case as an example. *Antitrust Bulletin*, **46**, 115 33.

Levinson, R., Romaine, R., and Salop, S. (2001). The flawed fragmentation critique of structural remedies in the Microsoft case. *Antitrust Bulletin*, **46**, 135 62.

Liebowitz, S. (April 30, 1999). Breaking Windows: Esti mating the cost of breaking up Microsoft Windows. Working paper, Association for Competitive Technol ogy and the ASCII Group Inc.

Rule, C. (1999). The proper standard for identifying monopolistic practices: Why Microsoft hasn't violated the antitrust laws. *Connecticut Law Review*, **31**, 1387.

US Circuit Court, District of Columbia (2001). *US* v. *Microsoft Corp.*, 253 F.3d 34 (D.C.Cir. 2001). Washington, DC: US Government Printing Office.

US District Court, District of Columbia (2000). *US* v. *Microsoft Corp.*, Civil Action No. 98-1233 Court's Final Judgement. Washington, DC: US Government Printing Office.

US District Court, District of Columbia (2001). *US* v. *Microsoft Corp.*, Civil Action No. 98-1232 (CKK) Proposed Final Judgement. Washington, DC: US Government Printing Office.

minimum efficient scale

Robert E. McAuliffe

The minimum efficient scale (MES) represents the smallest output level for a firm at which long run average costs are at a minimum (*see* LONG RUN COST CURVES). If the long run average cost curve were U shaped and continuous, then the MES firm size would be unique. However, statistical estimates of cost curves for various industries suggest that long run average cost curves are L shaped where there are significant ECONOMIES OF SCALE at low levels of output which are exhausted relatively quickly, then average costs remain constant (see Johnston, 1960; Scherer, Beckenstein, and Kaufer, 1975). This means that the MES represents a lower bound on firm size but not necessarily an upper bound.

Estimates of the MES have also been obtained using engineering surveys and survivor studies. With engineering surveys, industrial engineers and other experts provide information concerning the expected changes in costs as the scale of operations increases, and from this the MES is determined. Recognizing the difficulties in estimating the MES, Stigler (1958) suggested that those firms which survived in the competitive environment should be the most efficient (*see* EFFICIENCY). Survivor studies examine the changes in the number of firms in different size classes over time to determine the optimum size plant (or, as Stigler suggested, optimum range of sizes). In the SHORT RUN firms may not be operating at the optimal scale, so reliable survivor estimates must be based on industries in long run EQUILIBRIUM. Unlike engineering studies these studies use data from operating firms, but those firm sizes that survive could have done so through anticompetitive behavior or because of BARRIERS TO ENTRY, which would not reflect efficiency. Nevertheless, survivor studies tend to confirm the results obtained from the engineering and statistical cost studies: there appears to be a wide range of optimum plant sizes, suggesting a range of output levels where long run average costs are constant.

Knowledge of the plant level MES in an industry is important to understand the feasible number of firms that could operate in the industry. Scherer et al. (1975) concluded from their estimates of MES that actual concentration ratios in US industries are much higher than required by the estimated minimum efficient scale. If correct, these estimates indicate that antitrust policies which would break up large firms might not cause inefficiency. However, ECONOMIES OF SCOPE are also important in industry and estimates of economies of scale and the MES may fail to detect these additional causes of larger firm size, so any policy actions must be carefully considered. See Gold (1981) for a critical survey of the theoretical issues regarding firm size.

The behavior of costs for plants operating at less than MES is also important. If costs increase significantly when plants are smaller than the MES, then the disadvantages of small size are much greater and this could deter ENTRY into the industry.

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Industrial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Gold, B. (1981). Changing perspectives on size, scale, and returns: An interpretive survey. *Journal of Economic Literature*, **19**, 5 33.

Johnston, J. (1960). *Statistical Cost Analysis*. New York: McGraw-Hill.

Scherer, F. M., Beckenstein, A., and Kaufer, E. (1975). *The Economics of Multi Plant Operation: An International Comparisons Study*. Cambridge, MA: Harvard University Press.

Stigler, G. J. (1958). The economies of scale. *Journal of Law and Economics*, **1**, 54 71.

monopolistic competition

Kostas Axarloglu

Monopolistic competition is a MARKET STRUC TURE characterized by a large number of firms where every firm has some MARKET POWER with respect to its products. The concept was introduced by Edward Chamberlin (1933) to study deviations in terms of prices and the number of firms in the market from the perfectly competitive market structure (*see* PERFECT COMPETITION).

The basic market characteristics of a mono politically competitive industry are the following:

- there is a large number of firms, and every firm has a small market share;
- every firm sells a differentiated product (*see* PRODUCT DIFFERENTIATION) from its competitors, and therefore it has some market power (downward sloping market demand schedule);
- there is free ENTRY and EXIT of companies into and from the industry in the long run, which usually leads to lower market share and zero ECONOMIC PROFIT;
- because of the large number of firms in the industry, each firm does not consider the reactions of its rivals in its own decisions (*see* OLIGOPOLY).

The market structure of monopolistic competi tion is suitable to study several different ques tions which cannot be easily addressed in the context of other market structures. First, since every firm in the industry sells a different var iety, the level of PRODUCT VARIETY in the market can be assessed along with its social wel fare implications (*see* OPTIMAL VARIETY). Second, this market structure allows economists to explore the reasoning behind companies' de cisions on the type, design, and selection of the varieties (brands) they offer in the market. Finally, the implications of brand selection on companies' pricing policies can also be assessed.

In analyzing the monopolistically competitive market structure, two major families of models have been developed. **Chamberlinian models**, sometimes also called the *representative consumer* models, and **Hotelling type models**, also known as *location* or *address* models.

CHAMBERLINIAN MODELS

Introduced by Chamberlin (1933) and extended by Dixit and Stiglitz (1977), the Chamberlinian model is a representative consumer model with a large number of firms, each one of which offers a distinct product variety in the market. In the model, consumers have an insatiable desire for product variety and perceive every brand as an equally good substitute for every other brand available in the market. The ELASTICITY of substitution among different brands is exogen ously given and constant, and is not related to market entry or exit. In other words, the repre sentative consumer does not have an "ideal" variety, i.e., a variety he prefers the most. Firms produce one variety each and compete only in terms of prices, so they do not compete in the design and the particular characteristics of their products.

The number of different varieties in the market is equal to the number of firms in the industry. Firms will enter the market until all profitable opportunities are exhausted (zero eco nomic profits). Overall, the model predicts that monopolistically competitive industries will have a large number of varieties when there are low FIXED COSTS in production, and/or a low elasticity of substitution among different var ieties, and/or a large market demand.

However, the free entry of firms has some significant welfare implications on the economy (see the entry on optimal variety). Overall, these models predict that the product variety provided in the market can either be greater or less than optimal because of two contradicting forces. On the one hand, a firm will enter the market with a new brand if it believes entry will be profitable, but the new brand may take business from other firms already in the market. Since the entrant does not worry about the negative effect it will have on established firms, there tends to be too much entry and too many varieties. Offset ting this force is the fact that society may benefit more from a new brand than it costs to produce, but if the firm cannot practice PRICE DISCRIMINATION, it cannot capture all the benefits to society from introducing a new

brand and this force tends to reduce the number of brands.

The main criticism against these models, aside from the fact that they are based on a representative consumer, is the fact that they cannot explain how companies choose the attributes of their brands. There is no basis for a theory of *product choice* and *product design*. Also, the assumption that the representative consumer does not have an "ideal" variety is rather simplistic and not very appealing. Finally, the fact that the degree of substitution among different varieties is exogenously given, and is independent of the number of varieties in the market, does not allow economists to study the implications of the level of product variety on companies' pricing policies.

Recently, there have been some efforts to improve the basic Chamberlinian model, while preserving some of its fundamental characteristics. Perloff and Salop (1985) introduce a model where every consumer places a different relative value on each available brand. The consumer's preferences, along with the price of a brand, results in CONSUMER SURPLUS, and the consumer buys the brand which gives the highest surplus (the "best buy"). However, in their structure, all brands compete symmetrically with all the others (a Chamberlinian element).

HOTELLING-TYPE MODELS

In this family of models, product variety is due to the large number of consumers in the market with diverse preferences. In fact, it is usually assumed that consumers have an "ideal" variety for every product they consume. In other words, consumers can be "located" along the product characteristics space, a space where each point represents a different product variety. The consumer's location in this space indicates her ideal variety.

Introduced by Hotelling (1929), the model assumes that consumers are uniformly distributed along the product characteristics space and each one has an ideal product variety. The product characteristics space is linear (a straight line) and bounded, and can be considered equivalent to the market for the product. There are also many firms in the market, and each one produces one brand of the product. Firms decide about the particular attributes of these brands by choosing the location of their brands along the product characteristics space. Brands compete with only their neighboring products on either side of their location in the product characteristics space. Hence, the degree of substitution among different brands is not the same for all brands.

Based on this structure, Hotelling showed that in a case of a *duopoly* (a market with just two firms), when companies do not compete in terms of price, it is optimal for them to locate next to each other in the middle of the product characteristics space. In other words, firms choose varieties which are very close SUBSTITUTES, since that way they can achieve maximum market share. The reason for this is that consumers will purchase from the store that is closest to them in the market. If consumers were located along a one mile stretch of road, the first firm in this market would want to locate in the middle – one half mile down the road – because this location allows the firm to attract the greatest number of consumers. Given this, the next firm would want to locate next to the first because it will also attract the greatest number of consumers, and since the firms do not compete on price, only their location matters to customers. This result, known as *minimum product differentiation*, depends on specific assumptions of the model, as will be discussed later. High market demand, low fixed costs (weak ECONOMIES OF SCALE), and weak substitution among different varieties lead to high product variety in the market. Overall, this model predicts a larger number of varieties in the market than the Chamberlinian models do, because there is weaker COMPETITION among brands, since varieties compete only against their neighbors and not against all the other varieties in the market.

As mentioned, Hotelling's result of minimum product differentiation is a special one. It depends on the assumptions that companies do not compete in terms of price and that the market is bounded. Generalizing Hotelling's results, Eaton and Lipsey (1975) show that in a case of a large number of firms, there is an EQUILIBRIUM where firms locate in pairs along the product characteristics space with some space between the pairs. In other words, there is some *clustering* of varieties in the market.

Other researchers derive Hotelling's result of minimum product differentiation by assuming instead that consumers are not uniformly dis tributed along the product characteristics space but are clustered around certain points (in some cases in the middle of the product characteristics space). In other words, consumers tend to prefer some varieties relative to others. Then, com panies have incentives to locate in those areas of high demand (*thick markets*). This is one of the causes of geographic concentration of economic activity in large metropolitan areas (*spatial ag glomeration*).

Finally, Lancaster (1979) introduces a model where consumers perceive products as bundles of characteristics and have preferences over dif ferent collections of characteristics, not neces sarily over individual products. Therefore, by combining different products in their consump tion, they can end up with the set of characteris tics and qualities they find most desirable. Called the *characteristics approach*, this model allows formalization of different situations such as whether products can or cannot be combined in consumption, or whether such combinations preserve the characteristics of the separate prod ucts.

MONOPOLISTIC COMPETITION AND PRICE MARKUPS

Recent research has studied how the entry and exit of firms in a monopolistically competitive market influences the pricing policy of com panies, and especially their price markups (*see* MARKUP PRICING). In the context of the Chamberlinian models, price markups are con stant and not related to entry and exit because the degree of substitution among different var ieties is constant and not influenced by the number of brands in the market. In terms of Hotelling type models though, price markups are inversely related to the entry of firms since market entry implies a larger number of brands, stronger competition, a higher price elasticity of demand, and lower price markups. The conclu sion is that periods of high market demand and therefore high market entry (e.g., an economic boom or a seasonal increase in market demand during Christmas) are associated with lower price markups. Weitzman (1982) and others pre sent theoretical justification for this result, while

Barsky and Warner (1995) offer some empirical support.

Bibliography

Barsky, R. B. and Warner, E. J. (1995). The timing and magnitude of retail store markdowns: Evidence from weekends and holidays. *Quarterly Journal of Economics*, **110**, 321 52.

Chamberlin, E. H. (1933). *The Theory of Monopolistic Competition*. Cambridge, MA: Harvard University Press.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic com petition and optimum product diversity. *American Eco nomic Review*, **67**, 297 308.

Eaton, B. C. and Lipsey, R. G. (1975). The principle of minimum differentiation reconsidered: Some new de velopments in the theory of spatial competition. *Review of Economic Studies*, **42**, 27 49.

Hotelling, H. (1929). Stability in competition. *Economic Journal*, **39**, 41 57.

Lancaster, K. J. (1979). *Variety, Equity, and Efficiency*. New York: Columbia University Press.

Perloff, J. M. and Salop, S. C. (1985). Equilibrium with product differentiation. *Review of Economic Studies*, **52**, 107 20.

Salop, S. C. (1979). Monopolistic competition with out side goods. *Bell Journal of Economics*, **10**, 141 56.

Weitzman, M. (1982). Increasing returns and the founda tion of unemployment theory. *Economic Journal*, **92**, 787 804.

monopoly

Robert E. McAuliffe

A monopoly exists when there is a single seller of a product in the industry. As a consequence, the monopolist's demand curve is the same as the industry demand curve and is thus downward sloping (*see* DEMAND CURVES). A monopolist must be protected by BARRIERS TO ENTRY to remain a single seller while earning ECONOMIC PROFIT. The monopolist may be protected by licensing requirements, patents, its own STRA TEGIC BEHAVIOR, or ECONOMIES OF SCALE which prevent ENTRY into the industry.

Since the industry demand curve facing the monopolist is the *average revenue* curve, when it is decreasing the MARGINAL REVENUE curve must lie below it. To maximize profits the monopolist will produce until marginal rev enue equals MARGINAL COST, a condition all

profit maximizing firms must fulfill regardless of MARKET STRUCTURE (*see* PROFIT MAXI MIZATION). If the monopolist sells a single, nondurable product and market demand is

$$P = f(Q) \tag{1}$$

where P is the market price of the product, Q is the quantity sold, and $f(Q)$ is the inverse DEMAND FUNCTION, then the marginal rev enue for the monopolist is

$$MR = \frac{\Delta TR}{\Delta Q} = P + Q \times \frac{\Delta P}{\Delta Q} < P \tag{2}$$

where ΔTR is the change in TOTAL REVENUE and ΔQ is the change in the quantity sold (see Martin, 1994). The marginal revenue curve slopes downward because, in the absence of PRICE DISCRIMINATION, the monopolist must lower the price to all consumers of the product in order to sell an additional (marginal) unit, and this is why the third term on the right in equation (2) above is less than P. Multiplying and dividing the next to last term in equation (2) by P yields

$$MR = P \times \left(1 + \frac{Q\Delta P}{P\Delta Q}\right) = P\left(1 + \frac{1}{e_p}\right) \tag{3}$$

where e_p is the price ELASTICITY of market demand for the product. Since the elasticity of demand is always negative, the marginal revenue in equation (3) above will be negative whenever market demand is inelastic (when $|e_p| < 1$). This means that the monopolist could increase rev enues (and profits) by selling fewer units and raising price, so an unregulated, profit maximiz ing monopolist should never operate on the in elastic portion of its demand curve. Setting marginal revenue in equation (3) above equal to marginal cost (MC) yields the profit maximizing markup for the monopolist (*see* MARKUP PRICING):

$$MC - P = \frac{P}{e_p}$$
$$\frac{P - MC}{P} = -\frac{1}{e_p} \tag{4}$$

Thus the optimal markup for a monopolist is inversely related to the absolute value of the elasticity of demand. The term on the left hand side is the LERNER INDEX of monopoly power and shows that the monopolist will charge a higher price markup over marginal cost when consumers have fewer SUBSTITUTES (market demand is less elastic).

THE SOCIAL COSTS OF MONOPOLY

In the preceding discussion it was shown that a monopolist will set its price above marginal cost, while in perfectly competitive markets price equals marginal cost (*see* PERFECT COMPETI TION). This means that some consumers who are willing to pay the cost to society of producing this product do not receive it, and this is called DEADWEIGHT LOSS. Consider an example where, for simplicity, there are no FIXED COSTS in production, so that the long run mar ginal cost (LRMC) is equal to long run average cost (LRAC) in figure 1.

To maximize profits the monopolist will pro duce Q_m units of output (where MR = MC) and the market price for Q_m units will be P_m. If this industry were perfectly competitive it would produce until $P = MC$, which would be Q_c units sold at price P_c above. Compared with a perfectly competitive industry, the monopolist produces less output and charges a higher price. In figure 1, the rectangle P_m,M,B,P_c is the mon opolist's profit, which is a transfer from con sumers (who lose this amount in CONSUMER SURPLUS) to the monopolist. Since this transfer makes consumers worse off and the monopolist better off, the welfare consequences cannot be judged on the grounds of Pareto optimality (*see* PARETO OPTIMAL ALLOCATION). However, the triangle M,B,C is lost consumer surplus that is not gained by the monopolist. This is the deadweight (efficiency) loss to society from a monopoly. Once the monopolist has earned its profits, hypothetically it should be willing to produce additional units for those consumers who are willing to pay the marginal cost of pro duction (that is, those consumers along the demand curve between points M and C). This production would make consumers (and there fore society) better off and leave the monopolist no worse off, so it is a more efficient (Pareto optimal) allocation. Since the monopolist does not produce and sell these additional units, this loss in allocative efficiency is one of the costs to society from monopoly.
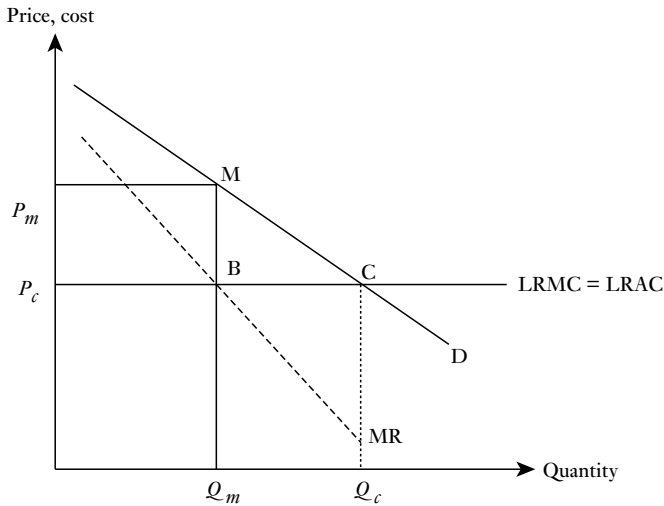
Figure 1   Profit-maximizing price and output under monopoly

Another potential cost of monopoly is what Leibenstein (1966) termed X inefficiency (*see* X EFFICIENCY), the failure to minimize production costs. Without the pressure of competition, a monopolist has less incentive to be efficient and may waste resources in production. Therefore the monopolist's costs would not be equal to those achievable under perfect competition, and so the social costs of monopoly would include the monopolist's higher costs of production in addition to the deadweight loss.

Potential monopoly profits may encourage firms or individuals to expend resources at tempting to acquire or maintain a monopoly. Such behavior is called RENT SEEKING and if these expenditures do not create benefits to society, then the effort to monopolize is costly to society. Posner (1975) argued that if firms compete for these monopoly profits, they would expend resources until the net expected profit was zero, and this meant that the social cost of monopoly was all of the monopolist's profits. Although this represents an extreme upper bound, some resources are wasted in rent seeking behavior and this increases the costs to society from a monopoly (see Tirole, 1988).

An interesting issue considered by Arrow (1962) was the effect of market structure on the incentives to innovate. In his model, firms could innovate to reduce their marginal costs of production, and he examined the returns from this innovation under perfect competition and monopoly. Arrow found that the monopolist had *less* incentive to innovate because it was already earning monopoly profits and the innovation would replace its existing position. On the other hand, a competitive firm could become a monopoly and would earn higher profits. Therefore monopolies may also be costly to society in terms of dynamic efficiency: these firms may fail to innovate or introduce new products. However, if the monopolist is threatened by entry, it has a greater strategic incentive to innovate and preempt potential entrants because in doing so it will remain a monopolist, while an entrant will have to share the market with the monopolist (see Tirole, 1988, for a survey).

Monopolies can be beneficial to society under certain conditions. When there are economies of scale, a single firm may be able to serve the market more efficiently than several firms (*see* NATURAL MONOPOLY). Governments grant monopoly licenses to firms and individuals through patents in an effort to encourage innovation. In this case, rent seeking behavior has socially desirable benefits which society promotes.

Bibliography

Arrow, K. J. (1962). Economic welfare and the allocation of resources for inventions. In R. Nelson (ed.), *The Rate and Direction of Inventive Activity*. Princeton, NJ: Princeton University Press.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Leibenstein, H. (1966). Allocative efficiency vs. "X-efficiency." *American Economic Review*, **56**, 392 415.

Martin, S. (1994). *Industrial Economics*, 2nd edn. New York: Macmillan.

Posner, R. (1975). The social costs of monopoly and regulation. *Journal of Political Economy*, **83**, 807 27.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

monopsony

Lidija Polutnik

Monopsony exists when one firm is the only buyer of an input or product. The classic example of a monopsony is a one company town where the only employer in the town is a textile mill or a coal mine. In purchases of inputs the monopsonist can derive its power from other factors as well. For example, purchases of "house brand" goods where suppliers supply most of their output to a single retailer provide the buyer with a significant amount of monopsony power (also buyer power).

How many inputs should the monopsonist hire? What should be their pay? In a competitive labor market, each employer takes the wage rate as given, and the MARGINAL COST of hiring one more worker is simply the wage rate. In this case the marginal cost of the input curve is horizontal and coincides with its supply curve. The supply curve facing the monopsonist is far different. A single buyer of labor in a textile mill town faces the entire labor supply of the town. While the competitive firm can hire all the labor it wants at the going wage, the monopsonist cannot. The monopsonist must pay successively higher wage rates to attract additional workers and the higher wage rate applies to all the workers previously hired. As a result the marginal cost of input (labor) curve lies above its supply curve. A profit maximizing monopsonist

hires workers up to the point where the MARGINAL REVENUE product as defined by the derived demand for labor of the last worker hired equals the marginal input cost. The wage rate is determined by the height of the labor supply curve at the level of employment. The monopsonist restricts hiring in order to pay a lower price for labor or other inputs. For example, in a monopsonistic labor market the firm pays a lower wage rate than would prevail in the case of a competitive market, because workers lack alternative sources of employment and specialized inputs in general can be sold to one or, at best, a limited number of users. A gap between the derived demand of an input and the supply curve reflects the loss in EFFICIENCY in a monopsony.

A monopsony, like a monopoly, confronts the long run problem of EXIT and ENTRY. If the single buyer succeeds in depressing the input price below the competitive price, it may not be feasible for a specialized input to exit immediately. However, in the long run suppliers will not replace specialized equipment when it becomes obsolete and the incentive to produce inputs that are limited to one or a few uses will decrease. A producer that is concerned about the long run relation with its supplier may be better off in not exercising its monopsony power in the SHORT RUN (Carlton and Perloff, 2000).

Perry (1978) demonstrates that a powerful incentive for BACKWARD INTEGRATION is provided by higher profits than can be earned if a monopsony integrates backward. If the monopsonist purchases its supplier, then it does not need to pay a higher input price on all the units used if it chooses to expand an input usage. The marginal input cost schedule becomes irrelevant and the integrated firm chooses to hire the input up to where the derived demand of input equals the market supply curve of the input. Complete backward integration in the input markets results in the allocatively efficient outcome of a competitive input market.

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Industrial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

Perry, M. K. (1978). Vertical integration: The monopsony case. *American Economic Review*, **68**, 562 70.

moral hazard

Alexandra Bernasek

Moral hazard is an incentive problem that arises in cases where the actions of individuals cannot be observed and contracted upon, creating ASYMMETRIC INFORMATION among individual parties to a transaction. Moral hazard arises commonly in insurance markets, financial markets, and labor markets. It arises in situations involving cooperative effort by two or more people, and in the context of principal–agent relationships (*see* PRINCIPAL–AGENT PROBLEM). In some cases of moral hazard the actions of individuals are unobservable, but in many cases the problem is that the costs of observability are prohibitively high. Those costs are described as "monitoring costs." The nature of transactions characterized by moral hazard is such that individuals do not have incentives to behave in ways that lead to Pareto efficient outcomes (*see* PARETO OPTIMAL ALLOCATION). Solving the problem of moral hazard involves designing incentive CONTRACTS that reduce monitoring costs and combine risk sharing with the creation of appropriate incentives.

Insurance markets provide a good illustration of the moral hazard problem; if people can insure themselves against certain risks they are less likely to act with an appropriate level of care. Consider the example of auto theft. If the probability of a theft occurring depends on the actions of the car owner (e.g., where the individual parks, whether or not they lock their car doors, etc.), then the insurance company faces an incentive problem; they want the car owner to take actions that minimize the probability of theft. In this case, full insurance will not be optimal since the insurance company will want the car owner's wealth to depend on his or her actions, thus creating incentives for the car owner to take the proper amount of care (e.g., this can be used to explain deductibles). The amount of insurance an individual can purchase

at actuarially fair rates is effectively rationed in the presence of moral hazard. An early article establishing these results is Spence and Zeckhauser (1971).

Holmstrom (1979) has shown that in the presence of moral hazard, Pareto optimal risk sharing in the context of a principal–agent problem is generally not possible because it won't induce the agents to take actions that are desired by the principal. A second best solution to the problem is possible which trades off some risk sharing benefits for the provision of incentives for the desired behavior. One solution is to spend resources on monitoring agents' actions and use this information in the contract. Holmstrom shows that any additional information about the agent's actions, no matter how imperfect, can be used to improve the welfare of both the principal and the agent.

Shavell (1979) also explores what Pareto optimal fee schedules would look like in agency relationships characterized by moral hazard. He finds that the characteristics of these fee schedules are related to both the principal's and the agent's attitudes toward risk (*see* RISK AVERSION). He discusses the implications of his results for four examples of principal–agent relationships: strict liability versus negligence standards in the control of stochastic EXTERNALITIES, insurance, the lawyer–client relationship, and the relationship between STOCKHOLDERS and managers of a firm.

The analysis of moral hazard in the principal–agent context extends beyond the principal–single agent to a multi agent setting. Two features of a multi agent setting that are not present in a single agent setting are free riding and COMPETITION. Holmstrom (1982) explores these as a way of better understanding the organizational design of firms. His focus is on "moral hazard in teams," where a team refers to a group of individuals who are organized in a way that productive inputs into the firm are related. He then examines the organization of production in this context when the agents' inputs are imperfectly observed. He finds that the free rider problem can be resolved if ownership and labor are to some extent separated, and that relative performance evaluation, such as peer production averages, can reduce moral hazard costs. The latter finding is one

explanation for executive incentive packages that base rewards on comparisons with peer firms (firms within the same industry).

Varian (1990) explores several issues related to group incentives in the context of a real world credit institution: the Grameen Bank in Bangla desh. One of the issues Varian explores is the use of groups of borrowers to reduce monitoring costs by the bank as a way of overcoming the moral hazard problem inherent in the credit relationship. The Grameen Bank creates incen tives for borrowers to monitor one another by requiring them to form groups of five from among their peers and accept joint responsibility for repaying all group members' loans. Because their access to loans is tied to the repayment performance of the group, agents have an incen tive to monitor their fellow group members to make sure they are taking actions that are con sistent with repaying their loans. Stiglitz (1990) also models this use of "peer monitoring" as a solution to the moral hazard problem in credit markets.

In an early article on the relationship between agency costs and the ownership structure of the firm, Jensen and Meckling (1976) examine the problem of moral hazard in the context of the separation of ownership and control in the modern corporation. There is divergence be tween the interests of owners of the firm and managers. A moral hazard problem arises be cause of the difficulty owners have observing the actions of managers and thus writing com plete contracts which require them to act in the owners' best interests. Real resources will there fore be allocated to monitoring activities by the owners (shareholders). In their definition of agency costs, Jensen and Meckling include the monitoring expenditures of shareholders: the costs associated with their attempts to measure and observe the actions of managers, as well as the costs of their efforts to control the actions of managers. Examples of these activities are audits, formal control systems, budget restric tions, and incentive compensation schemes.

Grossman and Hart (1980) examine the role of takeover bids in creating incentives for efficient management of the "stock market corporation." They assume significant costs in ensuring that managers and directors of corporations act in the interests of the shareholders. They model the stock market corporation as a common property resource (see PUBLIC GOODS PROBLEM) and explain the deviation between potential benefits and actual benefits of collective action in terms of the problem of moral hazard. The extent of this deviation will depend upon the extent of the unpredictability associated with takeover bids. They argue that the higher the threat of a raid, the more efficient current management of a cor poration will be. Shareholders will thus weigh the costs and benefits of creating the possible threat of a takeover bid as a way of monitoring the actions of managers.

The role of moral hazard in explaining the organization of production within firms has been an important area of research. Alchian and Demsetz (1972) examine information prob lems associated with the organization of produc tion to explain how "team production" induces the contractual process of firms. Team produc tion involves a cooperative activity, the essence of which is that the contribution of individual cooperating inputs to output cannot be identi fied. Their central question is one of incentives: how can members of a team be rewarded and induced to work efficiently? They conclude that monitoring to reduce shirking can be achieved more efficiently in a firm than through market bilateral negotiations among inputs. Mirrlees (1976) also explores issues of optimal payment schedules and organizational structure in cases where the performance of individuals in produc tion can only be imperfectly observed. His con clusion that IMPERFECT INFORMATION binds the organization together is consistent with the conclusions of Alchian and Demsetz.

Attempts to study the effects of moral hazard in the labor market have been important for explaining disequilibrium phenomena such as involuntary unemployment. Firms may pay above market clearing wages in order to prevent workers from shirking when their effort cannot be directly observed. These "efficiency wages" may increase worker effort for two reasons. One is the threat of unemployment that is created by the above market wages, and another is im proved worker morale. The efficiency wage is the wage at which labor costs are minimized. Efficiency wage theory has also been used to explain the observed downward rigidity of wages, and layoffs. For a survey of efficiency

wage models of the labor market, see Akerlof and Yellen (1986).

See also *asymmetric information; imperfect information; principal–agent problem*

Bibliography

Akerlof, G. A. and Yellen, J. L. (eds.) (1986). *Efficiency Wage Models of the Labor Market*. Cambridge: Cambridge University Press.

Alchian, A. and Demsetz, H. (1972). Production, information costs, and economic organization. *American Economic Review*, **62**, 777 95.

Grossman, S. and Hart, O. (1980). Takeover bids, the free rider problem, and the theory of the corporation. *Bell Journal of Economics*, **11**, 42 64.

Holmstrom, B. (1979). Moral hazard and observability. *Bell Journal of Economics*, **10**, 74 91.

Holmstrom, B. (1982). Moral hazard in teams. *Bell Journal of Economics*, **13**, 324 40.

Jensen, M. and Meckling, W. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, **3**, 305 60.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Mirrlees, J. A. (1976). The optimal structure of incentives and authority within an organization. *Bell Journal of Economics*, **6**, 105 31.

Shavell, S. (1979). Risk sharing and incentives in the principal and agent relationship. *Bell Journal of Economics*, **10**, 55 73.

Spence, A. M. and Zeckhauser, R. (1971). Insurance, information and individual action. *American Economic Review*, **61**, 380 7.

Stiglitz, J. (1990). Peer monitoring and credit markets. *World Bank Economic Review*, **4**, 351 66.

Varian, H. R. (1990). Monitoring agents with other agents. *Journal of Institutional and Theoretical Economics*, **146**, 153 74.

multicollinearity

Alastair McFarlane

Multicollinearity refers to the situation where the explanatory variables of a LINEAR REGRESSION are highly correlated with each other. This phenomenon is especially common in TIME SERIES DATA because of the presence of lagged variables and common time trends among explanatory variables. Perfect multicollinearity exists when an explanatory variable is an exact linear combination of other explanatory variables, making it mechanically impossible to calculate regression coefficients. This extreme case is unlikely unless the researcher has constructed a poorly specified model.

It is difficult to disentangle the separate effects of explanatory variables on the explained variable when multicollinearity is present. Coefficients may have the wrong sign or an implausible magnitude and small changes in the data can produce wide swings in the estimated coefficients. When combined with high VARIANCE in the error term and low variance in the explanatory variables, a high degree of multicollinearity will lead to high standard errors of the coefficients and low t statistics (*see* STANDARD ERRORS OF THE COEFFICIENT; T STATISTIC). However, multicollinearity does not adversely affect the predictive power of the regression model as a whole.

One method of detecting multicollinearity is to examine the correlation matrix for high correlation among the explanatory variables. A signal that multicollinearity is problematic is a high R^2 and low t statistics. More formal measures of the degree of multicollinearity are the condition number, the variance inflation factor, and Theil's measure (see Maddala, 2001).

A remedy is to increase the sample by adding more data which, by providing additional information, will help to lower variances. Since obtaining more data is usually difficult, researchers may resort to ad hoc techniques such as ridge regression, principal components regression, or omitting the offending variables. These ad hoc techniques have been criticized for producing biased or meaningless estimates. In such cases it may be advisable to accept multicollinearity and interpret the regression coefficients with caution.

Bibliography

Greene, W. H. (2002). *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.

Gujarati, D. N. (2002). *Basic Econometrics*, 4th edn. New York: McGraw-Hill Irwin.

Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.

multi–unit auctions

Brett Katzman

These are more elaborate AUCTIONS used to sell multiple units of a single good. Multi unit auctions have been used for decades by the US Treasury to sell notes, bills, and bonds, are used every day on eBay, and have garnered much attention due to their success in the sale of PCS spectrum rights.

There are four primary ways of auctioning multiple units. A method used for years by the US Treasury is the discriminatory (or pay your bid) auction, where bidders submit sealed bids. Those submitting the highest bids for units being sold are awarded the objects and must pay the (perhaps different and therefore discriminatory) amount of their specific bid(s). The US Treasury has recently begun experimenting with uniform price auctions where bidders also submit sealed bids. Those submitting the highest bids are awarded the items, but each winner pays the same price per unit, which is determined by the market clearing bid. The uniform bid is similar to the single unit second price auction, so much so that it is commonly (and falsely) considered the multi unit analog thereof. The actual analog to a second price auction is the Vickrey auction (named after Nobel Laureate William Vickrey), which has a much more complex payment rule. Finally, perhaps the simplest method for auctioning multiple units is to sell them sequentially using single unit auctions.

Bibliography

Cramton, P. and Ausubel, L. (2002). Demand reduction and inefficiency in multi-unit auctions. Working paper, University of Maryland.

McAfee, R. P. (2002). *Competitive Solutions: The Strategist's Toolkit*. Princeton, NJ: Princeton University Press.

N

natural monopoly

Robert E. McAuliffe

A natural monopoly is defined by Sharkey (1982) and Panzar (1989) as a market which can be supplied more efficiently (at lower cost) by a single firm than by two or more firms. The conventional treatment in textbooks suggests that ECONOMIES OF SCALE are required for a natural monopoly, but this need not be the case. For example, consider a single product firm with a U shaped average cost curve which has a minimum at 1 million units of output. If market demand were 1.1 million units and average costs did not rise too quickly, it would be more efficient for the single firm to produce this output than to have two firms in the market, and yet average costs would be rising for this single firm at an output level of 1.1 million units. Economies of scale are sufficient but not necessary for a single product natural monopoly. It should also be noted that there could be economies at the firm level in a multiplant operation that could justify a natural monopoly even if economies of scale were exhausted at the level of the plant.

When a firm produces multiple products, economies of scale are no longer sufficient for a firm to be a natural monopoly. In the multiple product case, ECONOMIES OF SCOPE and of scale are necessary for a natural monopoly along with cost complementarities (see Sharkey, 1982; Panzar, 1989). Even when a natural monopoly exists, Demsetz (1968) suggested that potential competitors should be willing to supply the market and this will constrain the monopolist's MARKET POWER so that regulation may not be necessary. Using Bell system data, Evans and Heckman (1984) tested to determine whether one firm could produce telecommunications output in the US at lower cost than two

firms over the period 1958–77. They found that for qualifying data points, costs would be lower if *two* firms produced that output. In other words, the Bell system did not satisfy the necessary cost requirements to be classified as a natural monopoly (see Baumol, Panzar, and Willig, 1982).

See also *contestable markets*

Bibliography

Baumol, W. J., Panzar, J. C., and Willig, R. D. (1982). *Contestable Markets and the Theory of Industry Structure*. San Diego: Harcourt Brace Jovanovich.

Demsetz, H. (1968). Why regulate utilities? *Journal of Law and Economics*, **11**, 55 65.

Evans, D. S. and Heckman, J. J. (1984). A test for subadditivity of the cost function with application to the Bell system. *American Economic Review*, **74**, 615 23.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Panzar, J. C. (1989). Technological determinants of firm and industry structure. In R. Schmalensee and R. D. Willig (eds.), *Handbook of Industrial Organization*. New York: North-Holland.

Sharkey, W. W. (1982). *The Theory of Natural Monopoly*. Cambridge: Cambridge University Press.

net present value

Robert E. McAuliffe

Used in CAPITAL BUDGETING decisions, a project's net present value (NPV) is the discounted CASH FLOW of benefits from that project. To calculate the NPV for a given investment, the net cash flow for each period must be calculated and discounted at the appropriate discount rate (the firm's COST OF

CAPITAL is frequently used). Therefore, if an asset which costs P_0 today will generate cash flows of C_i in each period i for T periods in the future, then the net present value is

$$\text{NPV} = \sum_{i\,0}^{T} \left[\frac{C_i}{(1 - r)^i} \right]$$

where C_0 will equal the cost of the asset, $-P_0$, representing a negative cash flow (an outlay) in the initial period. The cash flows from the in vestment must be calculated as after tax values and all of the effects of the investment on the firm's cash flows must be considered. One fea ture of the NPV method of evaluating projects is that it assumes cash flows can be reinvested and earn the firm's cost of capital; Brigham and Gapenski (1997) argue this is more appropriate than the reinvestment assumption using the INTERNAL RATE OF RETURN method. The NPV approach also provides a more consistent ranking of alternative investment decisions for capital budgeting than other methods (see Brig ham and Gapenski, 1997; Edwards, Kay, and Mayer, 1987).

Profit maximizing firms will maximize the PRESENT VALUE of the firm by choosing those projects which have the highest NPV per dollar invested. In the example above, the cost of capital used to discount the cash flows from the investment is assumed to be constant over the life of the investment, but this need not be the case and the NPV calculation can easily be modified to accommodate this change (see Brigham and Gapenski, 1997; Edwards et al., 1987; Shughart, Chappell, and Cottle, 1994).

If INFLATION is expected to occur during the life of the investment, then the estimated cash flows must be adjusted accordingly. The net present value above can be calculated using real (constant purchasing power) dollar values for both cash flows in the numerator and the discount rate in the denominator. Or managers can adjust estimated cash flows and the dis count rate for expected inflation in their calcu lations. In either case managers should be careful to note whether their firm's revenues adjust in the same manner as costs when infla tion occurs.

See also *nominal income and prices*

Bibliography

Brigham, E. F. and Gapenski, L. C. (1997). *Intermediate Financial Management*, 5th edn. New York: Inter national Thomson Publishing.

Edwards, J., Kay, J., and Mayer, C. (1987). *The Economic Analysis of Accounting Profitability*. Oxford: Oxford University Press.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Pub lishing.

Shughart, W. F., Chappell, W. F., and Cottle, R. L. (1994). *Modern Managerial Economics*. Cincinnati, OH: South-Western Publishing.

net present value criteria

Michael D. Curley

The rules for selecting value maximizing pro jects follow from the discounted CASH FLOW process known as net present value (NPV) an alysis. Conventionally stated, NPV analysis compares the present value of a project's initial investment (I_0) with the present value of project net cash flow (NCF) over its life, all discounted at an appropriate risk adjusted rate of return – essentially, the opportunity cost of capital:

$$\text{NPV} = -I_0 + \sum_{t\,1}^{n} \frac{\text{NCF}_t}{(1 + k)^t}$$

The criteria leads to the adoption of projects that are worth more to the firm and its owners than the projects cost. The firm accepts all independ ent projects whose NPV is greater than zero and rejects negative NPV projects. The NPV of a project represents the expected dollar addition to the market value of the firm adopting the project. In the case of interdependent projects, since NPV analysis compares present dollar values, value additivity holds and incremental analysis of project combinations is valid. In the extreme case of interdependent, mutually excl usive projects, one accepts that project with the highest NPV. With capital rationing – which is more often the case among divisions of a firm than external to the firm – NPV criteria can be

used to select the set of projects that maximizes firm value subject to a budget constraint. Add itionally, NPV criteria avoid all of the difficulties associated with other project selection methods such as the INTERNAL RATE OF RETURN, pay back period, and the accounting based average rate of return.

Sources of positive net present value projects lie in unexpected changes in product markets that affect the stream of income associated with projects, innovation, and market imperfec tions. The NPV rules, combined with capital market efficiency (*see* EFFICIENT MARKETS HYPOTHESIS), provide corporate managers with shareholder wealth maximizing decision criteria, thereby facilitating separation of owner ship from control without penalty.

network externalities

Maria Minitti

A network can be defined as a collection of interdependent agents. A network externality exists when the value of participating in the network increases as the number of agents in the network increases. Most often, in economics, network externalities exist when a product's value increases with the number of users. As a result, network externalities are also identified as "increasing returns to adoption" (Arthur, 1989) and are discussed in the context of the product's characteristics. As with any other product, the demand curve for a product exhibiting network externalities is downward sloping; it shifts upward, however, with increases in the number of units expected to be sold (*see* DEMAND CURVES).

Katz and Shapiro (1985) distinguish between direct and indirect network externalities. A net work externality is direct when the value to a specific user is directly affected by the physical use of other consumers. Direct externalities are found in most communications products such as telephone, email, computer operating systems, and the Internet. For example, the value of owning a phone is higher the larger is the number of other people who own a phone. A network externality is indirect when users of a product with a large market benefit from the development of complementary products that become abundant or inexpensive once the main good has gained enough users (*see* COMPLE MENTS). For example, the utility that computer operating systems provide for consumers depends on the number of compatible applica tions. Microsoft products have dominated the market since the mid 1980s. As a result, the majority of applications have been written to run on Microsoft. This has provided an indirect network externality to PC owners who adopted a Microsoft operating system.

Although very important in a variety of markets, network externalities are particularly significant in high tech industries. In the tele communications market, for example, in add ition to the quality of the product, consumers take into account the number of other consumers using the same provider before signing up for a service. In the market for software applications, Gandal (1994) tested whether the spreadsheet market exhibited a network externality and whether consumers place significant value on compatibility. His findings suggest the existence of network externalities and of the resulting willingness of consumers to pay a premium for the benefits of such externalities. It is not easy to explain many phenomena in high tech industries without considering network exter nalities. As a result, most literature on network externalities deals with issues of technology adoption and product compatibility, as in Arthur (1989), Katz and Shapiro (1986), and Econo mides (1996).

Because of their ability to generate increasing return to adoption, products exhibiting network externalities generate a self reinforcing effect such that the continuation and intensity of their use depend on past decisions of producers and consumers. This phenomenon is called path dependency. For example, the demand for tele phones is path dependent because it depends on the number of them sold earlier. That is, a par ticular sequence of past choices causes one prod uct to emerge as the standard for its class among all the competitors. A different sequence, how ever, could have caused an alternative outcome. This means that each person's choice contrib utes to the determination of the aggregate result, though not with equal weight: earlier choices have more "steering power." Early choices

push toward a specific path and later choices are guided into the same path.

LOCK IN is a consequence of path depend ency and an additional possible outcome of net work externalities. David (1985) provides the standard example, the QWERTY keyboard. At the time of mechanical typewriters, the dispos ition of letters on QWERTY reduced the inci dence of jamming; thus its technology was the least expensive and emerged as the industry standard. Later, as non jamming equipment was developed, touch typists found the Dvorak keyboard superior to QWERTY. Because of high SWITCHING COSTS, however, QWERTY remains the industry standard so far. Thus, the keyboard market is "locked in" to an inferior technology. Similarly, in the market for PC op erating systems mentioned above, several firms, such as IBM, tried to compete with Microsoft's Windows by introducing their own operating systems. Their products, however, lacked com patible applications to efficiently compete with Microsoft products. The lack of compatible applications made them less desirable, discour aged enough consumers from regarding them as a viable alternative to Windows, and, ultimately, prevented them from obtaining a sizable market share.

It is important noticing that, in the presence of network externalities, failure to survive in the market may be due not to the inferior quality of the products but to the lock in caused by the network externality. To the extent that the lock in prevents COMPETITION and the intro duction of new and possibly superior products, network externalities may be the source of ineffi ciency. From a strategic point of view, the emergence of QWERTY illustrates how net work externalities and the possibility of lock in create strong incentives for firms to act as first movers (see FIRST MOVER ADVANTAGES). Firms competing in an industry with network externalities have an incentive to use network power to create high switching costs and deter any competitor from entering the market so as to create monopolistic rents. Although favorable to the incumbent firm, the costs of this type of standardization are twofold. First, there is a loss of variety. Second, when only one technol ogy is available, some users are forced to select what is for them a more expensive technology.

From an economic modeling point of view, network externalities may generate unpredict ability and multiple equilibria. As mentioned above, products and phenomena exhibiting net work externalities possess a self reinforcing effect. As a result, the process of their adoption or development throughout an economy may not be predictable. In the case of a product, for example, predictability would imply the possi bility of forecasting exactly, for any type of insti tutional structure and economic condition, what the resulting level of usage will be. That is, it would permit ex ante deterministic predictions of the rate of adoption. Katz and Shapiro (1985), however, point out that consumers' expectations are the most crucial factor in markets where network externalities are present because differ ent expectations lead to different equilibria. Rohlfs (1974) shows that there are multiple equi libria at any given price in the communications industry. His work is based on interdependence of demand and assumes that a subscriber's utility increases as additional individuals subscribe to the same communications service. Rohlfs dem onstrates also that, for any service, there exists a minimum equilibrium user set with zero con sumers and a maximum equilibrium user set including the whole population and is able to show the critical mass needed in order to reach a desirable equilibrium. A comprehensive treat ment of the technical issues associated with net work externalities may be found in Arthur (1995).

Bibliography

Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *Economic Journal*, **99**, 116 31.

Arthur, W. B. (1995). *Increasing Returns and Path Depend ence in the Economy*. Ann Arbor: University of Mich igan Press.

David, P. A. (1985). CLIO and the economics of QWERTY. *American Economic Review*, **75**, 332 7.

Economides, N. (1996). Network externalities, comple mentarities, and invitation to enter. *European Journal of Political Economy*, **12**, 211 33.

Gandal, N. (1994). Hedonic price index for spreadsheets and an empirical test for network externalities. *Rand Journal of Economics*, **25**, 160 70.

Katz, L. M. and Shapiro, C. (1985). Network external ities, competition, and compatibility. *American Eco nomic Review*, **75**, 424 40.

Katz, L. M. and Shapiro, C. (1986). Technology adoption in the presence of network externalities. *Journal of Political Economy*, **94**, 822 41.

Rohlfs, J. (1974). A theory of interdependent demand for communications services. *Bell Journal of Economics*, **5**, 16 37.

nominal income and prices

Robert E. McAuliffe

Nominal values are simply current dollar values. Thus a consumer's nominal income is his or her current dollar income and the nominal price of a product is its current dollar price. Since eco nomic decisions should be based on real values rather than nominal values, the distinction bet ween nominal and REAL INCOME or nominal and REAL PRICES can matter in business deci sions. This distinction becomes much more im portant as the rate of INFLATION increases because the difference between real and nominal measures will grow with the rate of inflation. For example, if an industry's costs rise by 10 percent and consumers' nominal incomes also rise by 10 percent, then if firms raise their prices by 10 percent, the nominal price of the product will be 10 percent higher but the real price of the industry's product will not have changed (see Douglas, 1992).

Bibliography

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

normal goods

Nikolaos Vettas

A good is called normal if consumers choose to buy more of this good when their income in creases. More precisely, a consumer chooses a level of consumption for each good which is determined by the consumer's preferences and budget constraint (which is, in turn, determined by the consumer's income and the prices of goods). When a consumer's REAL INCOME in creases, then she can afford to consume more of each good. However, whether she chooses to do this or not depends on her preferences. If the consumption of a good increases following an increase in the consumer's real income, it is called normal (between the initial and the final income levels). If a good is not normal, it is called inferior (*see* INFERIOR GOODS). Of course, it is possible that a good is normal at some income levels and inferior at other levels. For example, as a consumer's income rises, the consumption of a given good may first increase and then decrease (such as preferences from a discount store).

An important property of normal goods is that their demand functions have a negative slope (*see* DEMAND FUNCTION). When the price of a normal good increases, both the INCOME EFFECT and the SUBSTITUTION EFFECT tend to decrease the quantity demanded.

Bibliography

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Pub-lishing.

O

objective of the firm

Robert E. McAuliffe

Although it is normally assumed that firms pursue PROFIT MAXIMIZATION as their primary goal, a variety of alternative assumptions has appeared in the literature. Conventional theory treats the firm as a single homogeneous unit with the single goal of maximizing profits, but Nelson and Winter (1982) argue that even if firms did attempt to maximize profits, they could only do so in familiar, repetitive situations. When a firm produces a new product or faces a new environment, it lacks the information required to maximize profits and must learn and adapt to the new situation. Herbert Simon (1959) suggested that firms cannot maximize profits because of the inevitable compromises and UNCERTAINTY which arise in corporate decision making and which lead to "satisficing" behavior rather than maximizing behavior. Under these circumstances, managers attempt to meet satisfactory goals for profits and sales rather than maximizing profits. But as Blaug (1992: 159) has noted, the satisficing hypothesis does not yet provide specific predictions to challenge or replace profit maximization.

Since many managers are concerned with market share, another possible goal for a firm's managers could be sales maximization (Baumol, 1962) subject to a minimum profit constraint. Oliver Williamson (1964) questioned why the managers of the firm should pursue maximum profits for the owners instead of their own self interest. He proposed a theory of managerial discretion where managers could use the resources of the firm for personal objectives rather than maximize profits. The assumptions and predictions of these theories will be compared with the assumption of profit maximization below.

SALES MAXIMIZATION

Baumol (1962, 1977) investigated the consequences of assuming that firms maximize sales (TOTAL REVENUE) subject to some minimum profit constraint. The minimum profit requirement is necessary because a firm could maximize sales but enter into BANKRUPTCY if profits were ignored. When the firm's goal is to maximize sales, Baumol found that the firm would spend more on ADVERTISING, service, and other demand increasing factors than would a profit maximizing firm. In addition, changes in FIXED COSTS or property taxes cause the sales maximizing firm to raise prices and decrease output. This contradicts the predicted behavior of the profit maximizing firm whose price and output are unchanged by a change in fixed costs, and businesses appear to raise their prices when fixed costs increase.

One interesting result discussed by Baumol (1977) is the multiproduct firm's choice of inputs and outputs under sales versus profit maximization. For a given level of costs, the sales maximizing firm will use the same combination of inputs and produce the same combination of outputs as the profit maximizing firm. This result occurs because profits are revenues minus costs and if the level of costs is fixed, maximizing profits must also maximize revenues. Therefore Baumol concludes that the different predictions from sales maximization must be due to differences in the total level of costs and revenues rather than simply a reallocation of a given level of costs. Intuitively, the sales maximizing firm reduces its profits from their maximum level to the minimum amount required, and this profit difference is a resource

which is allocated to different products in order to maximize sales. To increase sales the profit from a product is reduced (because the firm is producing beyond the point of maximum profits), so if sales are maximized, the firm must "spend" this reduction in profits where the sales increase will be highest. But just as profit maximizing behavior requires the firm to hire inputs to the point where all are equally productive, sales maximizing behavior requires the firm to increase output until the marginal increase in sales for one product (per dollar of profit sacrificed) is equal to the marginal increase in sales for another product (per dollar of profit lost).

Managerial Discretion

Since economists assume that individuals act in their own self interest, it is natural to expect that the managers of the firm (who are normally not the owners) would do the same. Williamson (1964) suggested that managers have goals other than profit maximization which they would pursue using the firm's resources. This is an example of the principal–agent problem, where the interests of the managers (the agents) hired by the owners (the principals) conflict with the interests of the owners. Managers might prefer larger staffs, bigger offices, and other cost increasing expenditures than would be consistent with maximizing profits.

The managerial discretion model predicts that the firm will vary the size of its staff and the level of perquisites directly with market conditions. Thus, if business conditions are poor the firm will reduce its staff and other non essential expenses and will approximate the profit maximizing level if market conditions worsen. In other circumstances, the level of staff employment will be greater than predicted for maximizing profits because of the managers' preference for staff. Furthermore, if fixed costs increase or a lump sum tax is imposed, the firm will reduce its output, staff employment, and its level of perquisites. In contrast, the profit maximizing firm would not change its price or output levels. Williamson also observes that when a new manager with a lower preference for staff is installed, significant decreases in staff employment can occur without reducing performance, a result

which is consistent with the managerial discretion theory.

Unfortunately, the satisficing, sales maximizing, and managerial discretion theories require knowledge about the decision making process of each firm and the preferences of the firm's managers for specific predictions. In this respect, these hypotheses are less general than the profit maximization assumption, although this does not mean that profit maximization is more accurate or appropriate. Rather, the alternatives must be better specified to replace the assumption of profit maximization in economic theory.

Bibliography

Baumol, W. J. (1962). On the theory of expansion of the firm. *American Economic Review*, **52**, 1078 87.
Baumol, W. J. (1977). *Economic Theory and Operations Analysis*. Englewood Cliffs, NJ: Prentice-Hall.
Blaug, M. (1992). *The Methodology of Economics*, 2nd edn. New York: Cambridge University Press.
Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.
Lesourne, J. (1992). *The Economics of Order and Disorder*. New York: Oxford University Press.
Nelson, R. R. and Winter, S. G. (1982). *An Evolutionary Theory of Economic Change*. Cambridge, MA: Harvard University Press.
Simon, H. A. (1959). Theories of decision making in economics. *American Economic Review*, **49**, 253 83.
Williamson, O. E. (1964). *The Economics of Discretionary Behavior: Managerial Objectives in a Theory of the Firm*. Englewood Cliffs, NJ: Prentice-Hall.

oligopoly

Sayed Ajaz Hussain

Competition within an oligopoly is characterized by strategic interdependence between firms, a feature lacking in other market structures (*see* market structure). For example, a monopolist has *complete* market power as its choices alone determine the price and output in the market, whereas a firm in perfect competition is a price taker and hence has *no* market power – it cannot by its choice of output influence the market price. In an oligopoly each firm is large enough to have an impact on price and output, but so do its rivals. The market

outcome therefore depends on the nature of interaction between firms.

Because the market outcome depends on the underlying nature of interaction between firms, an oligopoly can encompass a rich variety of outcomes: firms may fight price wars, collude, seek to differentiate their goods (*see* COLLUSION; PRODUCT DIFFERENTIATION), or choose to compete in multiple dimensions such as price, ADVERTISING, QUALITY, research and development (R&D), location, product attributes, etc. Because of the potential richness of outcomes in an oligopoly, even a rudimentary treatment of oligopoly behavior is impossible here. We therefore restrict our discussion to concentration measures, static models (COURNOT COMPETITION and Bertrand), dynamic models (especially collusion and punishment), and the DOMINANT FIRM fringe followers models of oligopoly. The central results are, firstly, that in static settings firms will not be able to collude, but can do so in a dynamic setting since there is scope for detection of, and punishment for, deviating (cheating) from collusive outcomes. Secondly, when a dominant firm sets prices, it must consider the effect of attracting entrants if the price is set too high, and of sacrificing profits if the price is set too low.

Two CONCENTRATION INDICES are used to measure market structure: the concentration ratio (CR) and the HERFINDAHL HIRSCHMAN INDEX (HHI). The CR is defined as the cumulative market share of the *top* (typically four or eight) firms. A loose oligopoly, using CR_4, is defined as 40 percent < CR_4 < 60 per cent. The HHI is the sum of squared market shares of *all* firms in an industry: an HHI of over 1,000 indicates an oligopoly. The Department of Commerce's Census of Manufacturers periodically releases CR figures, while the Department of Justice uses the HHI to approve or reject horizontal mergers (*see* HORIZONTAL MERGER GUIDELINES; MERGER GUIDELINES, 1992–7). Concentration indices are not without drawbacks for they rely on the number of firms as being indicative of market structure. Such measures do not capture the underlying STRATEGIC BEHAVIOR of firms. For example, a firm may possess a large share of the market and yet may set prices as if it is competitive so as to dissuade potential entrants, or firms in a Bertrand duopoly (see below) may exercise no market power.

We begin with two *static* models of oligopoly: Cournot and Bertrand competition. These models purposefully brush aside strategic behavior and provide a foundation for strategic competition. In its simplest form the Cournot model is a static (i.e., one period) game in which, by assumption (so as to bypass issues related with potential entrants), duopolistic firms choose outputs. Each firm produces a homogeneous good, and must choose output simultaneously.

The price in the market depends on the outputs of both firms. Therefore, the profit of each firm is a function of its and its rivals' output. Choice of output by any one firm depends on anticipating the choice of output by its rival, giving rise to interdependency. EQUILIBRIUM outputs are easily derived by noting that this is a static game of complete (but imperfect) information (*see* GAME THEORY; IMPERFECT INFORMATION). Each firm takes the other firm's output as given and computes its best response output, or reaction functions. Solving the reaction functions simultaneously yields the Nash equilibrium outputs, known as Cournot outputs. No firm will wish to unilaterally deviate from the Cournot output because it is the best choice given the output of the other firm.

There are several qualitative features worth noting about the Cournot model (see Church and Ware, 2000): firms will exercise some market power where the market price is greater than MARGINAL COST (recall that in perfect competition, price equals marginal cost), but at the same time, the price is less than MONOPOLY price. The market power of each firm is limited by the price ELASTICITY of market demand – the more elastic demand is, the lower is the markup over marginal cost. Firms with lower marginal cost will also have a larger share of the market. Relaxing the assumption of a duopoly and allowing for an arbitrary number of firms in the market implies that each firm's market share decreases, reducing its market power, emphasizing the importance of BARRIERS TO ENTRY, which has implications for strategy.

Since the Cournot price is higher than perfect competition and lower than monopoly, this suggests that monopoly profits are higher than Cournot profits – which is indeed the case.

This suggests that both firms have an interest in colluding by restricting output. The Cournot firms, however, face a PRISONER'S DILEMMA: both firms realize that colluding is in their interest, but collusion is not sustainable in a static setting. At the collusive output, each firm has an incentive to increase output and raise its profits. But so does the other firm, and there is common knowledge of this fact. Acting on their incentives alone leads firms (and prisoners) to reach a sub optimal outcome.

The Cournot model has an interesting result about the optimal number of firms in a market. While perfect competition implies that competition is socially desirable, this is not necessarily the case in a Cournot oligopoly if there are ECONOMIES OF SCALE present in production. As the number of firms rises, there is increased competition, lowering each firm's markup. Firms produce less output as they are "squeezed" in, which means that with increasing returns to scale, average cost of production rises, raising prices. Depending on which effect is greater, *restricting* the number of firms may actually *raise* welfare.

The Bertrand competition model shares the same assumptions as the Cournot model except that firms compete by choosing *prices* instead of output. Consumers purchase (one) good from the lower priced firm. Assuming ample capacity, it can be shown that the Nash equilibrium has each firm charging a price equal to the marginal cost – a "paradoxical" result for an oligopoly. Prices have fallen to marginal cost, which is the perfectly competitive outcome; firms in Bertrand competition have *no* market power. There is empirical evidence (Brander and Zhang, 1990) that the US airline industry is Bertrand competitive.

Relaxing the assumption of ample capacity and product homogeneity eliminates the "paradox." A lower price will result in greater output, which may hit the capacity constraint. And with product differentiation, a lower price can grab only so much market share. The Bertrand model exhibits a prisoner's dilemma result: both firms know it is in their interest to coordinate prices and charge monopoly prices, but each firm has an incentive to undercut its rival, cutting prices to equate marginal cost and reaching a suboptimal level of profits.

Which of the two models – Cournot and Bertrand – is applicable for markets with homogeneous goods, a fixed number of firms, and where strategic choices must be made simultaneously in a static setting? Kreps and Scheinkman (1983) have proposed the following scenario: firms follow a two stage game. In the first stage, firms invest in capacity equal to the output the firms will produce under Cournot competition. In the second stage, firms compete in prices so that the firms compete at the Cournot outputs chosen in stage 1. By *committing* to capacity (which takes time to change), firms ward off price wars in the second stage (recall the Bertrand paradox doesn't arise with limited capacity). This illustrates how choice of capacity can be used to affect the dynamics of competition in the SHORT RUN, and that the Cournot model is appropriate in situations where capacity constraints exist and where investments in capacity are "lumpy" owing to the large scale of investment. The Bertrand model seems appropriate when firms are not constrained by capacity (*see* CREDIBLE STRATEGIES; FIRST MOVER ADVANTAGES).

In the static models discussed above, the timeless setting implies that firms choose their strategies assuming that rivals will not react: there is no scope for strategic behavior to affect short run outcomes. A dynamic setting is needed for strategic competition.

A dynamic setting provides opportunities for firms to collude in order to raise profits above Cournot/Bertrand levels. Firms will be able to punish "cheaters" (i.e., firms which deviate from collusive outcomes) and thus are able to possibly sustain collusive levels. For collusion to be successful, firms must reach an agreement regarding price and output, and be able to enforce these agreements.

An effective collusive agreement depends on several factors: the elasticity of market demand, the relative number and size of firms colluding, and the extent of barriers to entry. Collusion itself may take two forms: explicit and tacit. Antitrust regulations in many countries (US: Section 1 of the SHERMAN ACT; European Union: Article 85 of the Treaty of Rome [*see* EU COMPETITION POLICY, 2004]; Canada: Section 45 of the Canadian Competition Act) have legal restrictions on explicit collusion.

One may argue that compliance with antitrust laws is a cost benefit issue: firms will trade off expected costs of collusion with the expected benefits. Thus the level of enforcement and speed and strength of punishment have a bearing on whether firms choose to comply with anti trust regulations (see Church and Ware, 2000, for additional discussion).

Tacit collusion occurs when firms signal their intention and anticipate rivals' reactions. A simple example is a firm in a Bertrand oligop oly raising its price, which is then matched by its rival – note how prices must be public for tacit collusion. Another example has been found in airlines through an elaborate scheme involving the Airline Tariff Publishing Company (*US* v. *Airline Tariff et al.*): airlines were found to col lude through start/end dates and fares for thou sands of city pairs. We now discuss the conditions for tacit collusion and punishment strategies to be successful.

There are several factors which complicate reaching an agreement on collusion (see Church and Ware, 2000): cost asymmetries (the lower cost firms will always have an incentive to under cut prices), product heterogeneity (firms must agree on price and non price factors), innovation (it is difficult to reach collusive agreements when the product is changing), incomplete informa tion (makes tacit collusion difficult since, say, prices may not be observable), UNCERTAINTY (firms may have varying expectations about future demand and costs), asymmetries in pref erences (when firms have different discount rates, the proclivity to match/undercut prices in tacit collusion varies), seller concentration (a greater size distribution and number of firms make collusive agreements difficult), and en forcement (if there is limited scope for enforce ment, collusive agreements are harder).

The scope of enforcement in sustaining collu sive agreements is a recent development in in dustrial organization that comes from Stigler (1964). According to this view, the ability to detect cheating, the speed of punishment, and the strength of punishment are key factors in sustaining collusive agreements. We now exam ine this in the context of a dynamic Cournot game, which also illustrates tacit collusion.

Consider a dynamic game, played over infin ite periods, in which two firms play the Cournot game each period. In a dynamic setting, firms can react to collusive attempts and punish cheaters, provided the punishment is credible and harsh. In a dynamic game, history matters: firms observe what their rivals have done and act accordingly. One outcome, independent of his tory and subgame perfect (see the entry on game theory), is to produce the Cournot output each period. Another strategy that permits collusion is tit for tat: produce the collusive output in each period provided the other firm has done the same; if a rival deviates from the collusive outcome, produce the Cournot output *forever*.

As in the discussion of dynamic games, espe cially subperfect Nash equilibrium, the question arises whether the punishment strategy of pro ducing the Cournot output forever is credible. It can be shown (see Church and Ware, 2000; Tirole, 1988) that there is a critical discount factor over which this strategy will sustain collu sion and below which it will not. The idea is that if firms value the future enough, there is a strong incentive to make the collusive profits in the long run rather than the windfall profits from deviat ing, followed by Cournot profits forever (Cour not punishment). Abreu (1986, 1988) has shown that there are additional (subgame perfect) pun ishments which are more severe than the Cour not punishment. There exist optimal penal codes which minimize a cheater's present value of profits: these punishment strategies satisfy both sustainability and credibility requirements – see Church and Ware (2000) for a simple exposition.

Dynamic models illustrate how the potential for punishing cheaters can sustain collusive agreements. There are several factors that affect the ability to detect cheating and the speed and strength of punishment. One is whether prices are public – if they are private it is difficult to check whether rivals are adhering to agreed upon prices. Even if prices are public, in the presence of, say, demand uncertainty, it is diffi cult to know whether prices are falling because of cheating or a fall in demand. In this case, firms can sustain collusive agreements by applying a "trigger price strategy": if prices have been at or above collusive levels, all firms agree to produce collusive output. But if the price falls below collusive levels (for whatever reason), all firms agree to produce Cournot output for a

pre agreed period of time. Note how even des pite a fall in price weak demand triggers a price war to serve as a warning to all, not to cheat.

For other factors, such as size of cartel (*see* CARTELS), lumpy orders, product differenti ation, EXCESS CAPACITY, elasticity of firm demand, multimarket contact, growth rates of markets, response times, and cost asymmetries, see Church and Ware (2000).

There are many practices which facilitate col lusion (see Church and Ware, 2000) by making it easy to check if some firm is cheating: some of these are the practice of open price policies (list prices are public), trade associations (permits information flows), PRICE LEADERSHIP, or Stackelberg oligopoly (where one firm moves first by announcing a "collusive" price which is then matched by followers), meet the competi tion clauses (where customers have a right to ask a current supplier to match another supplier's lower price), most favored nation clause (where a customer is entitled to the lowest price charged to any customer), multiproduct pricing formulas (which facilitates price comparisons across dis similar products), delivered price formulas (which specifies a fixed charge, thereby facilitat ing collusion, for delivery within an area), and resale price maintenance (in which a supplier specifies a minimum price for goods sold by retailer).

Finally, we discuss the dominant firm model of oligopoly. There are many industries with a dominant firm and "fringe" followers: for example, Intel in microchips and De Beers in diamonds. The dominant firm faces a tricky pricing issue: if it sets price too high, it makes large profits which eventually attracts other rivals and ultimately erodes its dominant pos ition. On the other hand, if the dominant firm sets price too low to ward off ENTRY, it sacri fices profits. This issue is known as "dynamic limit pricing," for the dominant firm has to choose a trajectory of prices over time which limits entry while allowing for healthy profits. Three factors matter in setting a trajectory of prices: the interest rates (high interest rates mean that the dominant firm will discount future profits heavily and therefore charge high prices now), the relative cost position of the dominant firm and followers, and the expected response of entrants to high prices. Firms can charge high prices, even with rivals entering, through product differentiation (branding; *see* BRAND NAME) and by innovation so the dom inant firm stays ahead of the fringe. Intel's ex perience suggests that branding and innovation worked for some time, but as rivals such as AMD have finally caught up, Intel can no longer charge premium prices. Instead, its strategy is to move out of microchips and into other areas such as servers and wi fi applications.

The discussion here has focused on the central issues in an oligopoly and several features – product differentiation, price leadership, empirics – have not been covered. Interested readers are referred to Church and Ware (2000) and the more advanced Tirole (1988).

Bibliography

Abreu, D. (1986). Extreme equilibria of oligopolistic supergames. *Journal of Economic Theory*, **39**, 191 225.

Abreu, D. (1988). On the theory of infinitely repeated games with discounting. *Econometrica*, **56**, 383 96.

Brander, J. A. and Zhang, A. (1990). Market conduct in the airline industry: An empirical investigation. *Rand Journal of Economics*, **21**, 567 83.

Carlton, D. W. and Perloff, J. M. (2000). *Modern Indus trial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Church, J. and Ware, R. (2000). *Industrial Organization: A Strategic Approach*. Boston: McGraw-Hill Irwin.

Kreps, D. M. and Scheinkman, J. A. (1983). Quantity precommitment and Bertrand competition yield Cour not outcomes. *Bell Journal of Economics*, **14**, 326 37.

Stigler, G. (1964). The theory of oligopoly. *Journal of Political Economy*, **72**, 44 61.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

opportunity costs

Robert E. McAuliffe

Opportunity costs are fundamental to econom ics, yet they are frequently overlooked in prac tice. The opportunity cost of any decision or choice made by consumers or firms is the value of the next best choice that was sacrificed. For example, consider a division within a firm that produces computer chips which are used by another division to assemble computers. The computer assembly division would prefer to

obtain the chips at the production cost of the manufacturing division (say, $100 per chip) to increase its profits. However, if those chips can be sold in the market at $250 per chip, then every internal sale of chips to the assembly division has an opportunity cost of $250 to the company, the revenue the company could earn if the chips were sold in the market (see Shughart, Chappell, and Cottle, 1994, for their discussion of internal transfer pricing between divisions).

Opportunity costs are also important for a company's long run planning. If the computer company in the example above earns a 10 per cent return on the capital it has invested in computer assembly but could earn 15 percent elsewhere, then this company should leave this industry in the long run because the opportunity cost of investment capital is not covered. Since this company could earn 15 percent investing its capital in another industry, in the long run it should EXIT the computer assembly industry.

The distinction between ECONOMIC PROFIT and ACCOUNTING PROFIT is based in part on the concept of opportunity costs. Accounting profits indicate whether the firm *is able* to stay in business, but economic profits indicate whether the firm *should* stay in business. Opportunity costs require more time and expense to measure relative to accounting costs which are provided in the course of normal operations. However, if opportunity costs are ignored, companies are likely to allocate their internal resources inefficiently. For example, suppose a company owns a building and does not pay rental or interest expenses on that building. Then accounting practices would not count rental fees as costs to the firm because they are not paid by the firm. But if the building could be leased or rented, there is an opportunity cost when the company uses the building and that cost should be considered in managerial decisions to maximize profits (*see* PROFIT MAXIMIZATION). If managers never consider alternative uses for the resources of the firm, how can they know that they have found the best use? The calculation of economic costs includes these opportunity costs and ensures that scarce resources are used where they are most needed.

See also *efficiency*

Bibliography

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Shughart, W. F., Chappell, W. F., and Cottle, R. L. (1994). *Modern Managerial Economics*. Cincinnati, OH: South-Western Publishing.

optimal variety

Kostas Axarloglou

To determine whether the variety offered in a market is optimal, the product variety produced in the market must be evaluated using social welfare considerations. A byproduct of the extensive research on PRODUCT VARIETY, optimal product variety is a welfare statement about the amount of product variety offered in the market, with respect to some level which is considered to be best from a social point of view. Obviously, the definition of the optimality criterion has a central role in the discussion, since the level of product variety provided in the market is evaluated with respect to this criterion.

When introducing a new variety in the market, a company has to compare the launching cost of the new product (usually costs related to R&D and necessary marketing cost for the new product) with the benefits from selling it. Also, by introducing a new variety, the company influences its competitors' and its own willingness to provide more variety in the future. In other words, a new brand creates an externality for competitors (*see* EXTERNALITIES). Specifically, there are three distinct externalities generated by the decision of the company to introduce a new product variety in the market. First, the company is not able to capture all the benefits from introducing a new product in the market (it does not capture the entire CONSUMER SURPLUS) unless it follows a special type of pricing policy called perfect PRICE DISCRIMINATION. Under these circumstances, firms have less incentive to offer new varieties, and the result might be an *underprovision* of product variety in the market. Second, by introducing a new variety in the

market, a company "steals" sales from brands offered by its competitors which are close SUB STITUTES to its own, and discourages them from introducing their own varieties in the market. This effect may lead to an *overprovision* of product variety in the market because con sumers were already served by existing brands (so they are not much better off), yet the new brand imposes costs on the established firms which the new firm does not consider. Finally, in the case of a firm which produces more than one variety, if it prices its products with a price markup above the cost of production (*see* MARKUP PRICING), the company will allow some market share for more additional varieties. Therefore, it makes it more profitable to offer more of these varieties in the market (*overprovision of product variety*).

Despite these factors, it is assumed that con sumers value product variety since they can con sume products that they prefer. Economists usually define the socially optimal level of prod uct variety as the one which maximizes the sum of consumer surplus and PRODUCER SURPLUS. Based on this criterion, they evaluate the level of product variety actually offered in the market.

By emphasizing the various different aspects associated with the introduction of new varieties in the market, economists reach different con clusions regarding the optimal level of product variety provided in the market. In his study of MONOPOLISTIC COMPETITION, Chamberlin (1933) concluded that the market produced more than the optimal level of product variety (overprovision). He obtained this result because he defined optimal product variety as that level which allowed companies to produce at the min imum of their average cost curve and therefore take full advantage of ECONOMIES OF SCALE. This conclusion was criticized by Dixit and Sti glitz (1977), who showed that the market could produce less than the optimal amount of product variety (underprovision) if the firms cannot cap ture all the benefits to consumers from introdu cing a new variety. Salop (1979) showed that market EQUILIBRIUM may result in more var iety than optimal, where the optimal variety is defined to minimize each firm's production costs and maximize consumer surplus while supplying every consumer in the market. Finally, Scherer (1979) attempted to test empirically whether the degree of product variety in the ready to eat breakfast cereals industry was socially optimal. He estimated the welfare gains from introducing a new variety and he compared them with the cost of launching a new variety in the market. He concluded that there was too much variety in the breakfast cereal product range.

Bibliography

Chamberlin, E. H. (1933). *The Theory of Monopolistic Competition*. Cambridge, MA: Harvard University Press.

Dixit, A. and Stiglitz, J. (1977). Monopolistic competition and optimum product diversity. *American Economic Review*, **67**, 297 308.

Lancaster, K. (1990). The economics of product variety: A survey. *Marketing Science*, **9**, 189 206.

Salop, S. C. (1979). Monopolistic competition with out side goods. *Bell Journal of Economics*, **10**, 141 56.

Scherer, F. M. (1979). The welfare economics of product variety: An application to the ready-to-eat cereals industry. *Journal of Industrial Economics*, **28**, 113 34.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

P

Pareto optimal allocation

Theofanis Tsoulouhas

A "feasible allocation" is Pareto optimal (or Pareto efficient) if there is no other feasible allocation that makes at least one of the agents in an economy strictly better off without making someone else worse off. A feasible allocation is weakly Pareto optimal if there is no other feasible allocation that makes all the agents in an economy strictly better off. Clearly, if an allocation is Pareto optimal, then it is weakly Pareto optimal as well, for if there is no allocation that can make at least one person better off without making someone else worse off, then there should be no allocation that can make everybody better off. An "allocation" is a specification of how much of each good each agent will receive. An allocation is "feasible" if the total amount of each good assigned to agents does not exceed the total amount available in the economy.

The Pareto optimality concept comes from the nineteenth century engineer and economist Vilfredo Pareto. Pareto optimality is a criterion of allocative EFFICIENCY in that it can be a solution concept for the problem of using and allocating goods and resources in an efficient way. Thus, if an allocation is Pareto optimal, goods are allocated in an efficient way; if it is not, there is room for mutually advantageous trade.

The use of the Pareto optimality criterion can be demonstrated by an Edgeworth box diagram. The Edgeworth box is named after the nineteenth century economist Francis Ysidro Edgeworth, who invented the analytical tools of INDIFFERENCE CURVES and "contract curves." The Edgeworth box diagram is used to characterize efficient allocations when there are two agents and two goods (see figure 1).

The dimensions of the box represent the total amounts of the two goods available in the economy. The lower left and the upper right corners are reserved as the origins for the consumption spaces of the two agents. Thus, any point in the box shows how much each agent will receive of the two goods (measured horizontally and vertically from his origin).

The curve from the lower left to the upper right corner in the Edgeworth box shows the set of Pareto optimal allocations and is called the "contract curve" because it shows the set of efficient allocations or "contracts." Along the contract curve, the indifference curves for the two agents are tangent to each other, so that their respective marginal rates of substitution are equal (*see* MARGINAL RATE OF SUBSTITUTION). Any allocation on the contract curve of the diagram, such as allocation A, is Pareto optimal and weakly Pareto optimal because it is not possible to make either one of the agents better off without making the other worse off. Allocations outside the contract curve where the indifference curves are not tangent to each other, such as allocation B, are not Pareto optimal or weakly Pareto optimal because trade between the two agents can lead to allocations on the contract curve that either make at least one of the agents better off without making the other worse off, or make both agents better off. Thus, when the marginal rates of substitution are not equal, the allocation of resources is not efficient, and there is room for mutually beneficial trade. Finally, if point B represents the "initial endowment" (which is the endowment of goods each agent has before any trade occurs), then the solid line is called the "core." The core is the subset of the weakly Pareto efficient allocations that are "individually rational," in the sense that each agent is better off by exchanging goods with the other
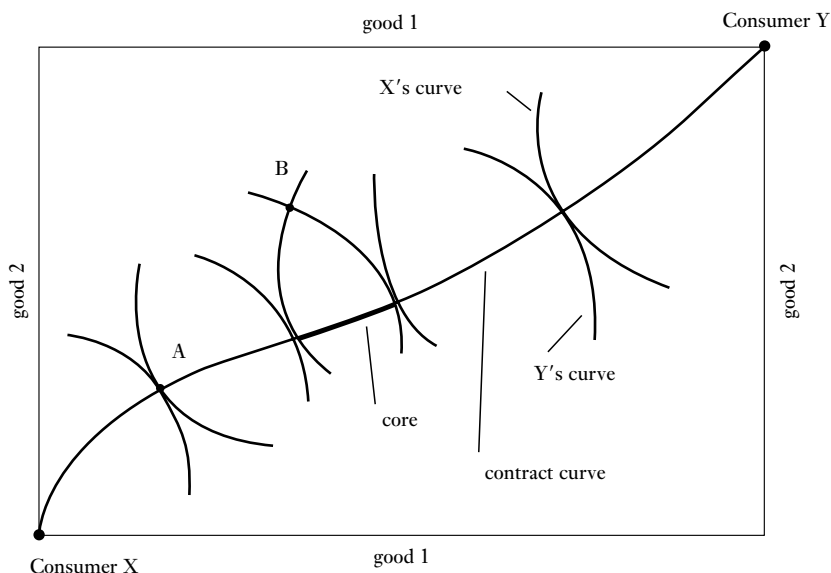
Figure 1   The Edgeworth box and the contract curve

agent than by simply consuming his endowment allocation.

See also *Coase theorem*

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Industrial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Cournot, A. A. (1927). *Researches into the Mathematical Principles of the Theory of Wealth*, 2nd edn. New York: Macmillan. (Originally published in French as *Recherches sur les principes mathématiques de la théorie des richesses*, 1838.)

Daughety, A. F. (ed.) (1988). *Cournot Oligopoly*. Cambridge: Cambridge University Press.

Vives, X. (1989). Cournot and the oligopoly problem. *European Economic Review*, **33**, 503 14.

pecuniary economies

Gilbert Becker

Pecuniary economies are cost savings resulting from large firm size which benefit the firm but not necessarily society as a whole. They are distinguished from real ECONOMIES OF SCALE in that the latter involves the reorganiza

tion of production in such a way that less of society's scarce resources are needed to produce a given level of output. As such, real economies not only benefit the firm by lowering its costs but also enable society on the whole to produce more goods.

Pecuniary economies, on the other hand, typ ically are cost savings which are generated from MARKET POWER. For example, sometimes a manufacturer, owing to its large size, is a large buyer of a raw material (or ADVERTISING ser vices, retail shelf space, CAPITAL, etc.) and is thus able to purchase that material at a lower price than some of its smaller rivals. Its lower costs, even if passed on to consumers, are ac quired at the expense of its suppliers. Here, no real resource saving occurs to society. Instead the end result of such economies is simply a wealth transfer from one firm (or group) to another.

Pecuniary economies can be beneficial to so ciety when they reflect economies of scale in the supply of inputs. For example, if a firm grows larger and increases its orders from suppliers, the suppliers may now be able to enjoy lower costs through economies of scale. Any subse quent lowering of price by suppliers to the pur chasing firm will create pecuniary economies for

that buyer. The purchasing firm does not achieve economies in production itself, but enjoys pecuniary economies because larger orders allow suppliers to be more efficient and lower their costs.

Pecuniary economies should by no means be discounted by the managers of a firm. Reductions in the COST OF CAPITAL, the price of electricity and other sources of energy, or the lowering of transportation costs all may significantly enhance a firm's profits. From a public policy point of view though, these gains should be treated differently. One example of this lies in the area of mergers. In evaluating the merits of a possible merger, the merging firms should consider both the real and pecuniary economies that may be achieved. Antitrust enforcement officials should consider only real economies in their evaluation, since pecuniary economies do not benefit society as a whole. In practice, mergers often bring both types of economies, and it may be difficult to differentiate between the two and identify the relative importance of each (see US Department of Justice and Federal Trade Commission, 1997).

See also *merger guidelines, 1992–7*

Bibliography

Shepherd, W. and Shepherd, J. (2004). *The Economics of Industrial Organization*, 5th edn. Long Grove: Waveland Press.
US Department of Justice and Federal Trade Commission (April 2, 1992; revised April 8, 1997). *Horizontal Merger Guidelines*. Washington, DC: US Government Printing Office.

perfect competition

Gilbert Becker

Perfect competition is one of the four BASIC MARKET STRUCTURES which exist in the economic classification system for markets. Although relatively few examples exist where markets completely fit the criteria for this market type, it plays an important role in economics and antitrust as it establishes a theoretical benchmark for the criteria needed to achieve socially optimal MARKET PERFORMANCE.

The criteria necessary for perfect competition are as follows:

1   The market contains a large number of buyers and sellers, each of which is small relative to the size of the whole market.
2   The product being produced by the rivals in this market is homogeneous and thereby lacking in features – ranging from product attributes (*see* PRODUCT ATTRIBUTES MODEL) to QUALITY differences to sales and service differences – by which each firm's product can be differentiated, in the eyes of consumers, from those of its rivals.
3   Firms in the market have knowledge of and access to the same technology for producing the product, and both buyers and sellers have perfect information as to other conditions in the market.
4   The ENTRY of new firms into the market may be described as being relatively easy to achieve in that no substantive BARRIERS TO ENTRY or EXIT deter new firms wishing to join the market.

Largely as a result of these criteria, managers of individual firms in this type of market find that they have no control over the price of their product as the combined forces of the market's supply and demand determine the market price for all firms. As such, perfectly competitive firms are said to be *price takers*. In addition, the ease of new entry, in theory, fosters a long run outcome in which price is high enough to cover all costs but does not offer any positive ECONOMIC PROFIT. Moreover, the threat of new entry promotes EFFICIENCY by forcing existing firms to minimize costs or be driven out of the market.

Baumol, Panzar, and Willig (1982) developed an alternative theoretical benchmark for examining market structure known as perfect contestability (*see* CONTESTABLE MARKETS), which holds as its predominant requirement costless entry and exit of firms. As perfect competition and this newer model both suffer from extreme requirements, industrial economists such as Shepherd (1982) call for the examination of criteria resulting in effective competition for evaluating market performance and public policy issues. Shepherd offers the requirements of:

1   reasonable parity among rivals;
2   sufficient numbers to preclude CARTELS;
3   behavior; and
4   the possibility of new entry.

Here, a group of ten firms, each having about 10 percent market shares, is seen as a reasonable first approximation, although situations wherein only half as many rivals exist may still be sufficient.

Markets most closely fitting the criteria for perfect competition often involve financial markets; commodities, such as agricultural goods; services, such as independent truckers; or the market for unskilled labor.

See also *oligopoly; monopolistic competition; mon opoly*

Bibliography

Baumol, W. J., Panzar, J. C., and Willig, R. D. (1982). *Contestable Markets and the Theory of Industry Struc ture*. San Diego: Harcourt Brace Jovanovich.
Maurice, C. and Thomas, C. (2002). *Managerial Econom ics*, 7th edn. Boston: McGraw-Hill Irwin.
Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.
Shepherd, W. (1982). Causes of increased competition in the U.S. economy, 1939 1980. *Review of Economics and Statistics*, **64**, 613 26.

pioneering brands

Robert E. McAuliffe

A pioneering brand is the first brand in a new market, or the first significantly different brand which creates a new market. When consumers have IMPERFECT INFORMATION about prod uct QUALITY, Schmalensee (1982) showed that the first brand in the market could enjoy signifi cant advantages over later entrants if the product performs satisfactorily. This advantage arises because the pioneering brand is able to set the standard by which all subsequent brands will be judged, and this can have lasting effects. Firms entering the market in later periods find it more difficult to convince consumers to incur the SEARCH COSTS necessary to learn about their brands and this creates a barrier to entry (*see* BARRIERS TO ENTRY).

In his study of the cigarette industry, Whitten (1979) found that the pioneering cigarette brands enjoyed substantially higher profits and market share relative to competing brands that followed. For example, the production of filter tip cigarettes opened a new market for female smokers and the early entrants into the national market, Viceroy and Winston, enjoyed consider able leads over their competitors.

Bibliography

Schmalensee, R. (1982). Product differentiation advan tages of pioneering brands. *American Economic Review*, **72**, 349 65.
Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.
Whitten, I. T. (1979). Brand performance in the cigarette industry and the advantages of early entry. Staff report to the US Federal Trade Commission.

predatory dumping

Kent A. Jones

Predatory dumping is the practice of inter national PRICE DISCRIMINATION (DUMPING) by an exporter with the intention or result of driving domestic producers in the targeted export market out of business. The predatory scenario is based on the assumption that an ex porter charges a high price in his home market and then subsidizes export sales in a foreign market until local COMPETITION is driven out, after which the exporter monopolizes the market and raises prices. In order for the strategy to be successful, the exporter must have a sufficiently "deep pocket" to underprice any rivals in the target market for a time sufficient to drive them out of the market, and then be able to prevent reentry or new ENTRY of rival firms (local or foreign) so that losses during the predatory stage can be recovered.

Strictly speaking, this predatory scenario, fo cusing on injury to competition itself, is ex tremely unlikely, since trade liberalization, reduced transportation costs, and increased competition in world markets have reduced the ability of firms to pursue price discrimination and create worldwide monopolies, especially in undifferentiated intermediate product markets,

which are the most likely targets for a dumping strategy. Often, however, the term "predatory" is used in a broader sense to describe injury to domestic producers in the importing country, who may lose market share or EXIT the market due to increased import competition. This concept of injury ignores any gains to consumers from the lower prices caused by dumping. Thus, antidumping laws treat all dumping as "predatory" if it results in injury to domestic import competing firms.

predatory pricing

Govind Hariharan

The practice of pricing products, sometimes even lower than costs, in order to drive out or coerce competitors or to prevent ENTRY of new firms is referred to as predatory pricing. Typically, credible threats are sufficient and need be carried out only once in a while to convince competitors. DUMPING products in a market with the purpose of driving out or acquiring rivals cheaply through lower prices is also considered to be a predatory pricing technique. Thus, larger firms and those with deeper pockets who also face potential gains from controlling the market (*see* MARKET POWER) are more likely to find it worthwhile to engage in such predatory pricing.

This practice is considered to be one form of PRICE DISCRIMINATION and a restriction of free COMPETITION, and typically faces a challenge in court under the antitrust laws (*see* ANTITRUST POLICY (US)). The dissolution of Standard Oil Company in 1911 is a case in point. The company increased its market share to over 90 percent of the oil refining industry and was accused of achieving this through acquiring its rivals cheaply and attempting to drive away the rest through threats of selling below cost. The Supreme Court determined in its ruling that this behavior violated the antitrust laws because the objective of the company in undertaking this practice was to monopolize the market (*see* MONOPOLY), and thus forced the breakup of the company into 30 independent firms. One study (McGee, 1958) that looked at whether

Standard Oil Company bought its competitors below their true value found that, in fact, Standard Oil had paid very handsomely for its acquisitions. Later decisions by the court, however, tended to also require proof that the company, having obtained control over the market, was likely to engage in acts that were detrimental to consumers.

Bibliography

McGee, J. (1958). Predatory price cutting: The Standard Oil (N.J.) case. *Journal of Law and Economics*, 1, 137 69.

Petersen, H. C. (1993). *Business and Government*. New York: HarperCollins.

present value

John Edmunds and Roberto Bonifaz

Cash flows which arrive at different times should not be compared until they are adjusted for the amount of time that elapses between each inflow (*see* CASH FLOW). For example, an immediate inflow of $1,000 is not the same as ten annual inflows of $100 each, although each is nominally equal to $1,000. The reasons they are not the same are:

1   profitable investment opportunities exist, which means that the $1,000 immediate inflow is better than the stream of ten $100 inflows (see OPPORTUNITY COST);
2   INFLATION exists, so the purchasing power of the two alternatives is not the same;
3   time preference exists, i.e., individuals will choose to have current consumption in preference to consumption later, unless they are compensated for deferring consumption.

These three reasons are related but conceptually distinct. All three of them, and one more besides, which is the risk of the cash flows, are embedded in the opportunity discount rate (ODR), which is the rate used to discount the cash flows.

The amounts that are coming in later time periods need to be discounted in order to be commensurate with an amount that is now

in hand. The process of discounting the cash flows that are to arrive in the future is called calculating the present value. A payment which is to arrive in one year's time, in order to be compared to a payment which is arriving today, must be divided by the amount $(1 + r)$, where r is the annual opportunity discount rate. This discounting is required because funds which are in hand today could earn a return over the course of the year of r. A payment which is arriving in two years' time, in similar fashion, needs to be divided by the amount $(1 + r)^2$ before it can be compared to a payment which is arriving today. The number which is calculated by adding the value of the payment arriving today, plus the discounted value of future payments, is called the present value.

In algebraic terms, if payments are CF_0, $CF_1, CF_2, \ldots, CF_n$ arriving today, one year from today, two years from today, and yearly until year n, then the present value (PV) of this flow of funds is calculated as:

$$PV = CF_0 + \frac{CF_1}{(1+r)^1} + \frac{CF_2}{(1+r)^2} + \ldots + \frac{CF_n}{(1+r)^n}$$
$$(1)$$

Or, more generally, if payments are received once a year until year n:

$$PV = \sum_{i\,0}^{n} \frac{CF_i}{(1+r)^i} \qquad (2)$$

It should also be noted that in the present value calculations above, the interest rate used to discount the cash flows, r, is assumed to be the same over the n years. For investment decisions, managers may want to consider the effects that variations in the discount rate would have on their NET PRESENT VALUE calculations.

Bibliography

Brigham, E. F. and Gapenski, L. C. (1997). *Intermediate Financial Management*, 5th edn. New York: International Thomson Publishing.

Edwards, J., Kay, J., and Mayer, C. (1987). *The Economic Analysis of Accounting Profitability*. Oxford: Oxford University Press.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

price discrimination

Kostas Axarloglou

Price discrimination is a specific pricing policy followed by companies with MARKET POWER, where different prices are charged for the same units of a product which are sold to different consumers, at the same or different periods of time. Companies use this pricing policy in an effort to increase their revenues and profits.

In market structures where companies have market power, the profit maximizing price for a product is given by the simple formula

$$P = \frac{1}{1 + \frac{1}{e_p}} MC,$$

where P is the price of the product, MC is the MARGINAL COST, and e_p is the price ELASTICITY of market demand, i.e., the sensitivity of market demand with respect to price. The company determines its price by adding a *price markup* to its marginal cost (the company's per unit cost), which is inversely related to e_p (*see* MARKUP PRICING). This optimal price markup applies when the company charges the same price, P, for all the units it sells. On the other hand, since consumers value the same product differently, they are willing to pay different prices for it (each has a different RESERVATION PRICE). If possible, it is beneficial for the company to charge a higher price to those consumers with a higher reservation price instead of charging the same price to everybody. This way, it manages to increase its revenues without increasing its cost of production and overall profits are enhanced. In other words, the company has strong incentives to *price discriminate* across different consumers and, as will be discussed later, even across different units purchased by the same consumer. This idea is at the core of the concept of price discrimination. Actually, price discrimination across different consumers allows the company to extract some or all of the surplus consumers receive from purchasing the product (*see* CONSUMER SURPLUS) and adds it to its revenues and profits.

For successful price discrimination the following conditions are necessary:

- *Non transferable products.* Products must not be easily transferable across consumers, or groups of consumers, who pay different prices for the same product. In other words, the TRANSACTIONS COSTS of transferring units of the product between different con sumers or groups of consumers must be rela tively high, so that the transaction is not profitable. On the contrary, if the transaction cost is low (*transferable commodities*), price discrimination is not effective, since those consumers who buy a product at a low price can sell (transfer) it to those who pay a higher price. In other words, the markets must be separate either geographically or by the identity of the consumers (*see* ARBI TRAGE).
- Consumers, or groups of consumers, must have different reservation prices for the same product. If consumers had the same price elasticity of demand, then the profit maxi mizing price the company should charge each consumer would be the same.
- The company must be able either to receive some information or design a method to extract information about consumers' reservation prices and willingness to pay. However, this is not easy, since consumers do not have incentives to reveal their true reservation price.

TYPES OF PRICE DISCRIMINATION

In economic literature, there is traditionally a distinction of three different types of price dis crimination.

First degree price discrimination. Sometimes called *perfect price discrimination*, this is the pricing policy where each consumer is charged a price equal to her reservation price for the product. Since consumers have different reser vation prices, they end up paying a different price for consuming the same product. It is very difficult and costly for any company to accurately determine the reservation price of each of its consumers, but this policy represents the maximum potential gain a company could extract. The service sector is the one where this type of pricing policy might be found, since services are not easily transferable products. Also, companies are able to more accurately

determine the reservation prices of their custom ers, although these estimates will be imprecise. For example, a law firm can charge different fees to different clients for the same type of legal advice, since it can extract some information about the willingness of its clients to pay the fees, based either on their income or, in the case of a company, on its sales or profits. Simi larly, a consulting company can follow the same type of pricing, since the services it provides are client specific, and therefore not transferable across clients.

Second degree price discrimination.   In this case, the company charges different prices for the same product based upon the quantities pur chased by consumers. Each consumer faces the same schedule of prices but will choose different quantities (and therefore pay different prices) based on their evaluation of the good. This form of price discrimination requires less infor mation about each consumer's reservation price and the company simply separates consumers into different groups according to how much they are willing to buy. For example, companies may offer discounts to consumers who buy the product in larger quantities, and by offering different quantities the firm can separate con sumers based on their willingness to buy the product (*see* BUNDLING).

Third degree price discrimination.   This is the most extensive form of price discrimination in practice. Specifically, the company charges dif ferent prices to different groups of consumers who appear to have different reservation prices. Obviously, the firm tends to charge a higher price to the group of consumers with the lower price elasticity of demand.

   There are many examples of this type of price discrimination. In the consumer goods industry, the use of coupons and rebates is widespread. A company charges the same price to all of its customers and then allows them to use discount coupons or rebates so they can buy its products at a lower price. The idea is that only those consumers who are sensitive to price (high price elasticity of demand), and therefore not willing to pay a lot, will clip, save, and use the coupons in buying the product. Obviously, this is an indirect and effective way to reveal con sumers' reservation prices. Also, senior citizen

or student discounts in entertainment or public transportation are examples of a third degree of price discrimination. Senior citizens and students usually have lower reservation prices for many of the products they consume. Companies then charge them a lower price than other consumers who do not fall in these two categories.

In the airline industry, all tickets for the same flight with the same airline do not cost the same amount. The ticket for business class costs more than for economy, in part because business travelers have a more inelastic demand for travel and cannot always make arrangements in advance. Tourists are much more price sensitive and are willing to order their tickets early. In addition, the airline companies offer better services to passengers traveling first class and this gives some justification for those who are willing to pay a premium for the first class cabin. This type of price discrimination is sometimes called *quality discrimination*.

There are other types of price discrimination which cannot be easily classified under one of the above major types.

Intertemporal price discrimination.   In this case, a company charges different prices to consumers who buy its product at different points in time. It is a pricing policy used extensively in the electronics industry when a new model is introduced in the market. The company initially sells the product at a rather high price for those consumers who are willing to pay the price and want to use it immediately, and later the price is decreased for those consumers with lower reservation prices (and as unit costs decline).

Finally, there are several other types of price discrimination. In the *two part tariff* case, consumers pay an up front fixed fee, which gives them the opportunity to consume a certain product, and then a per unit fee for the amount of units they actually purchase. Examples include amusement parks and various types of membership clubs. In the *peak load* price discrimination policy, companies charge a higher price when their demand peaks up and a lower price during the off peak period. Airline companies use it when they charge lower fares if passengers travel on Saturday evenings, or during off peak seasons. Movie theaters use it when they charge more for the evening show than the matinee.

Bibliography

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Publishing.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

Varian, H. R. (1989). Price discrimination. In R. Schmalensee and R. D. Willig (eds.), *Handbook of Industrial Organization*. New York: North-Holland.

price leadership

Robert E. McAuliffe

As Stigler (1968) observed, this term has been used in the literature with different interpretations. In the case of a DOMINANT FIRM, the dominant firm determines the industry price while the smaller producers act as price takers and sell at that price. However, the term has also been applied to those firms which are the first to announce their price changes and are typically followed by the other firms in the industry. In this case the "price leader" is not necessarily a dominant firm in the industry with substantial market share, and Stigler suggested that a firm might occupy this position if its prices correctly and rapidly reflected market conditions. Shughart (1990) argues that such a price leader may simply have better information or decision making skills than rival firms who simply follow the price change. When the leading firm's price changes reflect changes in costs or market conditions, it is called *barometric price leadership* because the leader's price acts as a barometer of the market environment. This form of price leadership is consistent with COMPETITION in the market and does not necessarily indicate COLLUSION between the firms.

The crucial question in price leadership models is: why should other firms follow the price changes announced by the leading firm? When competing firms are price takers, as in the dominant firm model, the other firms follow the price leader because they are so small relative to

the size of the market that they believe their output will not affect industry price. With a barometric price leader, rival firms take advan tage of what they recognize as the better infor mation or decision making of the leading firm. But in the case of a concentrated OLIGOPOLY, price leadership has been suggested as a method of tacit collusion (Markham, 1951) which helps firms coordinate their pricing. Price leadership has been suspected in the cigarette industry (Stigler, 1968), steel, the ready to eat cereals industry (Scherer and Ross, 1990), gasoline and coal industries (Stigler, 1968), but alternative explanations of observed pricing behavior are also plausible, such as MARKUP PRICING (see Shughart, 1990). Since collusive price leadership and barometric price leadership are difficult to distinguish empirically, Shughart (1990) argues that additional evidence is required to establish collusive behavior.

Bibliography

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Markham, J. W. (1951). The nature and significance of price leadership. *American Economic Review*, **41**, 891 905.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Publishing.

Scherer, F. and Ross, D. (1990). *Industrial Market Struc ture and Economic Performance*, 3rd edn. Boston: Houghton-Mifflin.

Shughart, W. F. (1990). *The Organization of Industry*. Boston: Richard D. Irwin.

Stigler, G. J. (1968). *The Organization of Industry*. Chicago: University of Chicago Press.

principal agent problem

Lidija Polutnik

The general problem of motivating one person or organization to act on behalf of another is known as the principal–agent problem. The principal–agent problem arises when the princi pal hires an agent to perform tasks on her behalf and the agent thereby influences the welfare of the principal. The principal–agent relationship provides a useful framework for analyzing situations in which there is ASYMMETRIC INFORMATION and when there is a need to design a contract (*see* CONTRACTS) or monitor the behavior of parties. MORAL HAZARD and ADVERSE SELECTION are also examples of the principal–agent problem.

For example, a typical firm is owned by shareholders (principals) who hire professional managers (agents) to run the company. The manager may be more interested in maximizing the firm's market share, size, and growth in order to provide her and her subordinates greater opportunities for promotion. Further more, managers may prefer to make investments whose payoffs come earlier rather than later, avoid risks, shirk, and otherwise fail to maximize the profits of the firm. Although economists commonly assume PROFIT MAXIMIZATION when we describe the decision making of a firm, the incentives of managers often differ from those of the shareholders and the efforts of the managers are impossible or too expensive to monitor. This would not be a problem if an enforceable contract could be drawn up that specified every duty of the manager and matched performance incentives to outcomes perfectly. Given that INCOMPLETE CONTRACTS occur frequently, due to bounded rationality, the firm's profit is less than the profit maximizing level and the difference is referred to as the residual loss.

Agency costs arise as principals try to ensure that the agents will act in the best interest of the principal. There are three types of agency costs: monitoring costs, bonding costs, and the residual loss of a principal (Jensen and Meckling, 1976). The principal–agent framework is employed to analyze the role of monitoring and bonding ac tivities in reducing the residual loss of the prin cipal. Monitoring costs occur when the principal employs resources to observe the efforts of the agent or creates incentives for the agent to undertake actions that are more likely to assure efficient use of resources within the firm. For example, an employee receiving a fixed salary

may be able to shirk on the job, thus the firm needs to develop a way of monitoring the performance, as well as the honesty, of the agent. Markets can also perform the monitoring function. In particular, the market for corporate control and the labor market for managers penalize managers who manage companies poorly. Thus, poor managers face a greater probability of unemployment and company takeovers, lower salaries, or a reputation for having brought a firm into BANKRUPTCY (Fama, 1980).

In order to guarantee performance, the agent may be required to post a bond. If the agent does not fulfill the terms of their agreement, then she must forfeit the bond. Edward Lazear (1979) showed that for workers who expect to make a career within an organization, it may be to the advantage of both parties to align the incentives of the employee and the employer by paying the employee below her MARGINAL PRODUCT early in the career and above her marginal product later in the career. This rising wage pattern is similar to a worker posting a bond to be collected later in her career. In this compensation scheme it becomes efficient for the employee to work diligently in order to avoid being fired before the deferred rewards can be collected (or the bond is forfeited). For this scheme to be effective, employees must find the firm's promise to be credible and the employer must provide for mandatory retirement. Similarly, stock options and bonuses that can be exercised by the manager only when she retires or leaves in a mutually agreeable way create incentives for managers to maximize long run profits of a firm. If the behavior of the manager is not acceptable to the owners of the firm, the manager faces reductions in the value of the deferred compensation.

Structuring the employee and managerial compensation packages is one of the ways in which the owners of a firm minimize the residual loss. In addition, owners of a firm who monitor the market price of their shares can thereby infer whether the agents are acting in their best interest; on the basis of that information alone they can buy and sell shares and limit the amount of residual losses imposed on them by the actions of the management. Hence, the power of STOCK HOLDERS to sell their stake in the firm promotes efficiency in the use of resources. In addition, a principal–agent problem is avoided when the owner of a firm also serves as the manager; in this case the firm eliminates the agency costs and any residual loss.

Bibliography

Fama, E. F. (1980). Agency problems and the theory of the firm. *Journal of Political Economy*, **88**, 272 84.

Jensen, M. C. and Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs, and ownership structure. *Journal of Financial Economics*, **3**, 305 60.

Laffont, J. J. and Martimort, D. (2001). *The Theory of Incentives: The Principal Agent Model*. Princeton, NJ: Princeton University Press.

Lazear, E. (1979). Why is there mandatory retirement? *Journal of Political Economy*, **87**, 1261 84.

Samuelson, F. W. and Marks, S. G. (2003). *Managerial Economics*, 4th edn. Danvers, MA: John Wiley.

prisoner's dilemma

Robert E. McAuliffe

The prisoner's dilemma is a classic illustration in GAME THEORY. Two suspects are arrested by the police and are charged with a minor crime. They are also suspected of a major crime but there is not enough evidence to convict them without a confession. The two prisoners are separated and are offered the following deal:

1   If one prisoner confesses to the major crime and implicates the other, he will serve 6 months in jail while the other will serve 6 years in jail.
2   If neither prisoner confesses to the major crime, each will be convicted of the minor offense and will serve 1 year in prison.
3   If both prisoners confess to the major crime, they will both serve 3 years in prison.

The "payoffs" to the two prisoners, Al and Bill, depend on the choice made by the other. These are shown in the matrix overleaf where Al's payoff is given in parentheses.

		Bill	
		Confess	Be silent
Al	Confess	(–3) –3	(–0.5) –6
	Be silent	(–6) –0.5	(–1) –1

The interesting feature of these payoffs is that the best strategy for each suspect individually is to confess. Consider Al's payoffs from choosing to confess compared with the payoffs from choosing to be silent. Whether Bill chooses to confess or be silent, Al's best strategy is to con fess, and the same result holds for Bill. Thus both prisoners are worse off pursuing their indi vidual self interest than if they agreed to be silent.

The strategy pair (confess, confess) is a *Nash equilibrium* because it is the best response for each player given the strategy of the other (*see* EQUILIBRIUM). That is, neither player can do better knowing the strategy chosen by the other. The prisoners would choose the same strategies if they used the minimax or MAXIMIN CRITER ION to evaluate their decisions (see Rasmusen, 1994).

If the players developed reputations to make the threat of punishment a credible threat, then the prisoner's dilemma can be avoided. Rasmu sen notes that reputation can be important in a variety of repeated games such as duopoly (where the strategies are to maintain the current price or cut the price), employer–worker rela tions (where the worker chooses to work hard or slack off), product QUALITY under IMPERFECT INFORMATION (where the firm chooses to pro duce a high quality product or a low quality product), and ENTRY deterrence (where the incumbent firm may retaliate aggressively or accommodate an entrant; *see* ACCOMMODA TION).

See also *adverse selection; signaling*

Bibliography

Gibbons, R. (1992). *Game Theory for Applied Economists*. Princeton, NJ: Princeton University Press.

Osborne, M. J. and Rubenstein, A. (1994). *A Course in Game Theory*. Cambridge, MA: MIT Press.

Poundstone, W. (1993). *Prisoner's Dilemma*. New York: Anchor Books.

Rasmusen, E. (1994). *Games and Information*, 2nd edn. Oxford: Blackwell.

producer surplus

Robert E. McAuliffe

Producer surplus is the difference between the TOTAL REVENUE received by the producers of a product and the minimum amount they would require to produce and sell the product. Since the supply curve measures the minimum price required by suppliers to produce and sell the product, any amount they receive above that minimum price would be the producers' sur plus. Thus the area above the market supply curve and below the market price measures the producer surplus in that market.

Producer surplus occurs because some factors of production are more productive than others, such as fertile land in agriculture or a prime location in a city. If these resources are more productive over time, they are more valuable to the owners and should command a higher market price, which is an economic rent (*see* RENTS). Since these resources have a higher value, if the firm's managers properly accounted for the opportunity cost of these resources, the firm's costs would be higher (*see* OPPORTUN ITY COSTS). Firms cannot earn ECONOMIC PROFIT in the long run when the industry is perfectly competitive, but they can earn eco nomic rents (*see* COMPETITION; PERFECT COMPETITION).

Producer surplus also plays a role in welfare analysis. Social welfare is frequently defined by economists as the gains from exchange and

production by consumers and producers. Maximizing social welfare then requires maximizing the sum of producer surplus and CONSUMER SURPLUS. One of the significant achievements of modern welfare economics was the demonstration that perfectly competitive markets in general EQUILIBRIUM can, in principle, maximize social welfare (Bator, 1957; Baumol, 1977; *see* PARETO OPTIMAL ALLOCATION).

Bibliography

Bator, F. M. (1957). The simple analytics of welfare maximization. *American Economic Review*, **47**, 22 59.

Baumol, W. J. (1977). *Economic Theory and Operations Analysis*. Englewood Cliffs, NJ: Prentice-Hall.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Publishing.

product attributes model

Kent A. Jones

Developed by Kelvin Lancaster (1971), the product attributes model sets out to explain consumer behavior as a process of choosing bundles of product characteristics or attributes inherent in goods and services, rather than simply choosing bundles of goods or services themselves. The basic assumption of the model is that the consumer's choice is based on maximizing utility from the product attributes subject to a BUDGET CONSTRAINT (*see* UTILITY MAXIMIZATION). The model is particularly useful in analyzing differentiated product markets, in which specific products that are SUBSTITUTES for each other are distinguished by their embodiment of a specific set of characteristics.

For purposes of exposition, a two dimensional graph reveals the model's main features (see Douglas, 1992, for textbook treatment and examples), and links it to the traditional budget constraint and indifference curve analysis of consumer behavior (*see* INDIFFERENCE CURVES). Figure 1 shows three specific products, each offering a specific amount of attribute X and attribute Y in constant proportions. Each

unit of product A contains X_a of attribute X and Y_a of attribute Y, for example. Similarly, each unit of products B and C offers the attribute bundles (X_b, Y_b) and (X_c, Y_c), respectively. The attributes could represent calories (X) and vitamin content (Y) for competing brands of soups, for example. While some attributes can be measured objectively in this way, it may also be useful to consider more subjective attributes, such as "atmosphere" and "quality of food" in distinguishing among restaurants, for example. Subjective attributes do, however, imply that the attribute content of a particular product may be determined largely by the perceptions of the individual consumer (see discussion of ADVERTISING below).

For a given budget constraint and set of prices for the products, the end points A, B, and C represent the limits of consumption along each attribute ray, and the line segment ABC defines the budget (or efficiency) frontier for the consumer. The consumer's choice is made by maximizing utility, as defined by the consumer's set of indifference curves, subject to the budget constraint. In this model, we interpret the slope of an indifference curve at a particular point (MARGINAL RATE OF SUBSTITUTION) as the rate at which the consumer is willing to trade off units of attribute Y for an additional unit of attribute X to remain at constant utility. Thus, the consumer's choice is influenced by his or her preference pattern in attribute space. As shown in figure 1, this consumer shows a strong preference for attribute Y and therefore chooses product A; a strong preference for attribute X would lead him or her to choose C. In this regard, the proliferation of differentiated products in a particular market can be explained as the result of the dispersion of tastes for various attributes among the population of consumers.

Note that the consumer will spend the entire budget on a single product A, B, or C if the highest indifference curve just touches the respective end point. If the highest indifference curve touches a point on the line segment between two attribute ray end points, then the consumer would choose to split consumption between the two adjacent products. If the product's consumption is indivisible (as in the case of an automobile or house), then such consumption splitting would be impossible, and the
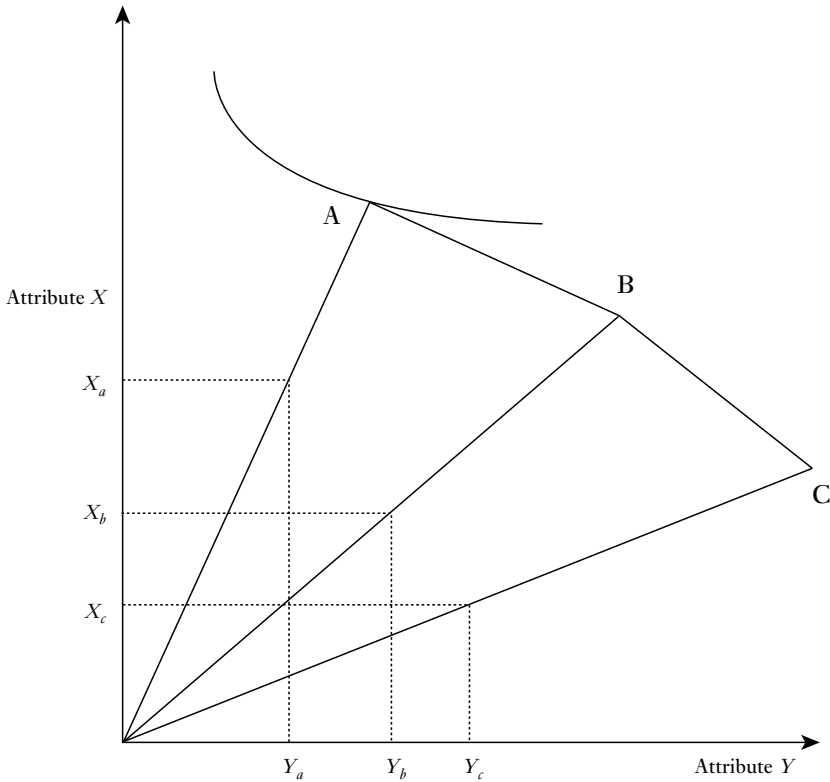
Figure 1   Product attributes and consumer choice

consumer's choice would be determined by the highest indifference curve that touches an end point (see Douglas, 1992).

The product attributes model also allows an analysis of STRATEGIC BEHAVIOR by firms and its effect on consumer choice. A decrease in the price of a product moves the end point further out along the ray, for example. Advertising can change the perception of the product in terms of attribute content and proportion (length and slope of the product ray) or the consumer's tastes for attributes (shape of individual consumers' indifference curves). Product positioning strategy may focus on opportunities based on "gaps" in the attribute space between rays, or may target existing products for direct COMPETITION.

Bibliography

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Lancaster, K. (1971). *Consumer Demand: A New Approach*. New York: Columbia University Press.

Salvatore, D. (2004). *Managerial Economics in a Global Economy*, 5th edn. Mason, OH: South-Western Publishing.

product differentiation

Robert E. McAuliffe

When firms in an industry sell products which are each distinct in the eyes of consumers, those products are said to be differentiated. Eaton and Lipsey (1989) suggested that differentiated products could be any set of closely related products in consumption and/or production. The degree of substitution between differentiated products varies and depends upon consumers' tastes in consumption, their information about other brands, and upon

ECONOMIES OF SCOPE in production. When firms produce differentiated products, each firm acquires some MARKET POWER and faces a downward sloping demand curve for its brand (*see* DEMAND CURVES). If consumers prefer particular brands relative to others, they have fewer SUBSTITUTES in consumption and so product differentiation reduces the ELASTI CITY OF RESIDUAL DEMAND for the firm's product(s). Caves and Williamson (1985) noted that product differentiation requires two condi tions: consumers must believe that differentiated goods within a product class are close substitutes for each other (relative to other products outside that class), and yet consumers must find the differentiated brands to be imperfect substitutes so that each firm faces a downward sloping demand curve. They found that IMPERFECT INFORMATION and complex product attributes (*see* PRODUCT ATTRIBUTES MODEL) com bined with FIXED COSTS were sufficient to ex plain product differentiation in US and Australian manufacturing industries.

Products may be differentiated *horizontally* or *vertically*. Horizontal differentiation occurs when consumers have diverse preferences and do not agree on which product is best. Vertical differentiation exists when all consumers may agree which product is best in terms of QUAL ITY but purchase different brands because of differences in prices or in their incomes. (See the entry on quality for more discussion of ver tical product differentiation.)

Theoretical analysis of product differentiation in economics has proceeded along two major lines: spatial or location models (sometimes re ferred to as address models) and representative consumer or Chamberlin models. Address models assume consumers have diverse prefer ences and that each consumer has a "most pre ferred" brand which she will purchase. In these models consumers have preferences defined over the product attributes which are embodied in varying proportions in each good (see Lancas ter, 1966; Archibald, Eaton, and Lipsey, 1986). In contrast, representative consumer models assume that consumers prefer diversity and pur chase each of the differentiated goods offered in the market (Dixit and Stiglitz, 1977). For example, people might enjoy eating dinner at different restaurants so if a new restaurant opened, it would share the market with the other establishments in the market. The address models treat the product space as continuous and provide insights into firms' product selec tion decisions, while in representative consumer models consumers have preferences for goods and so the product space is discrete. In represen tative consumer models each brand competes against all other brands in the market, whereas COMPETITION between brands is localized in the address models because any firm will have at most two competing neighbors on either side in a one dimensional market (see Archibald et al., 1986). Since most of the interesting issues con cerning product differentiation can be illustrated with address models, the discussion will focus on that class.

Consider a market one mile in length with N identical consumers distributed uniformly over that distance who will purchase one unit of the good if the total cost of the purchase is less than their RESERVATION PRICE, R, or nothing otherwise. When consumers must travel from their location to purchase a product, it is costly for them. If their transportation costs are linear, the total purchase cost is given by:

$$p + t \times \theta \qquad (1)$$

where p is the price of the good, t is the trans portation cost per unit of distance, and θ is the "distance" between the consumer and the firm. All else equal, consumers will purchase from the firm that is closest to them to minimize their total costs. The transportation cost may reflect the time and other travel costs required to pur chase the product (in physical space), or it could reflect the cost to the consumer of lost TOTAL UTILITY from purchasing a brand which is not exactly her most preferred brand (in product characteristics space). All else equal, a consumer would prefer to buy the brand closest to her ideal brand.

Where along this route would a firm want to locate? If the firm wants to be accessible to as many customers as possible, it should locate at the center of the market, one half mile from either end, since this location will minimize the average transportation cost to consumers in the market. Now suppose a second firm wants to enter this market and, for the moment, assume

that the price of the product in the market is fixed and will not change. In this case the second firm would also choose to locate at the center of the market because this location provides the firm with the greatest potential market share. This result led Hotelling (1929) to conclude that, in the absence of price competition, market competition will lead to *minimal product differentiation* where firms offer virtually identical products to appeal to as many consumers as possible. Although this result is not always an EQUILIBRIUM configuration (see Ireland, 1987; Tirole, 1988, and the sources cited there), it does explain the lack of diversity in broadcast television programming, political platforms, and other cases where price competition does not occur.

If price competition does occur, the firms will want to increase product differentiation to reduce the intensity of price competition. To illustrate, consider two firms, each located at opposite ends of the mile long market, and assume that the prices they charge are sufficiently low (or that consumer reservation prices, R, are sufficiently high) so that all consumers purchase one unit of the product. A consumer at location θ^* will be indifferent purchasing from either firm 1 or firm 2 when the "full" prices paid are equal:

$$R - p_1 + t \times \theta^* = R - p_2 - t \times (1 - \theta^*) \quad (2)$$

where p_1 is the price charged by firm 1 and p_2 is the price charged by firm 2. This indifferent consumer at location θ^* determines the market share of each firm, which can be found by solving equation (2) above for θ^*:

$$\theta^* = \frac{(p_2 - p_1 + t)}{2t} \quad (3)$$

When the two firms charge identical prices, the indifferent consumer is at half the distance and the firms share the market equally. If both firms have the same constant MARGINAL COST of production, c, it can be shown (see Tirole, 1988) that the profit maximizing price each firm will charge is equal to:

$$p_1 = p_2 = t + c \quad (4)$$

As transportation cost (or, equivalently, product differentiation) increases, each firm has more local monopoly power and competes less for sales, so equilibrium prices will rise. In fact, firms have incentives to pursue *maximal* differentiation to reduce price competition and will choose to locate at the extremes of the market when transportation costs are quadratic (see Tirole, 1988).

When firms compete in markets with differentiated products, there are forces which tend to increase the level of differentiation (and may possibly create a fragmented industry structure; *see* FRAGMENTED INDUSTRIES), and forces that work in the opposite direction toward *minimal* differentiation. As previously discussed, firms would prefer to increase differentiation to avoid price competition, all else equal. But as Tirole observed, firms will also want to locate near consumers and this reduces product differentiation. For example, if consumers were clustered at the center of the one mile market, we would expect firms to locate near the center of the market where the density of consumers was highest. This means that if consumers' tastes are not very different, the brands offered in the market should not be very different either. Another force reducing differentiation is positive EXTERNALITIES between firms which cause them to cluster together. These externalities arise when the costs of supplying firms in the industry fall when they share a common geographic location or when resources are more easily acquired (such as hiring computer professionals from nearby universities or competitors). Or, if firms share a geographic location, such as a shopping mall, they can reduce consumer SEARCH COSTS and raise their total demand even if price competition is intensified. Finally, Tirole notes that firms will tend to cluster when there is no price competition. It should also be noted that the amount of product differentiation supplied by a free market is unlikely to be the optimal level: unregulated markets may create too much or too little product variety (*see* OPTIMAL VARIETY). For example, in the one mile market described above, if the two firms choose their locations and then compete on price, they will choose to locate at the extremes of the market. However, since consumer demand is inelastic, the optimal locations should be those which minimize

transportation costs at one quarter and three quarters distance.

Salop (1979) developed a model where firms were located symmetrically around a circular market with free ENTRY and would costlessly relocate when entry occurred. He found that an increase in the fixed costs of production would lower the equilibrium number of brands and would cause prices to *fall*. If the size of the market increased or if consumer reservation prices rose, the equilibrium number of brands would increase.

Firms may also choose their locations strategically to create BARRIERS TO ENTRY in address models. Eaton and Lipsey (1989) show that firms could earn positive economic profits even when entry is "free" (*see* ECONOMIC PROFIT), and Schmalensee (1978) argued that firms in the ready to eat breakfast cereals in dustry prevented entry through PRODUCT PROLIFERATION. Both conclusions arise from strategic product positioning by incum bent firms. Consider a product space which is circular to avoid problems with end points (see Ireland, 1987; Tirole, 1988) and where con sumers are uniformly distributed around the circle. If the firms in the industry chose their locations strategically, they could locate in such a way that each firm earned economic profits and yet any firm which tried to enter the indus try would be unable to cover its fixed costs. For example, if there are product development costs which are SUNK COSTS once a firm enters the industry, then established firms could locate brands throughout the product space so that an entrant's best location would be between two existing brands. As long as the entrant's share of this area is less than its sunk costs, entry will be deterred, even though established firms earn economic profits. Schmalensee argued that there were brand specific sunk costs of ADVERTISING in the ready to eat breakfast cereals industry and that the estab lished firms offered so many brands that they literally crowded the product space and pre vented entry. Eaton and Lipsey (1989) note that in these models there may be economic profits even with free entry, the free entry in dustry equilibrium will not be unique, product variety will not be optimal, and monopoly power may persist as established firms make strategic decisions to deter entry.

Bibliography

Archibald, G. C., Eaton, B. C., and Lipsey, R. G. (1986). Address models of value theory. In J. E. Stiglitz and G. F. Mathewson (eds.), *New Developments in the Analysis of Market Structure*. Cambridge, MA: MIT Press.

Carlton, D. W. and Perloff, J. M. (2000). *Modern Indus trial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Caves, R. E. and Williamson, P. J. (1985). What is prod uct differentiation, really? *Journal of Industrial Econom ics*, **34**, 113 32.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic com petition and optimum product diversity. *American Eco nomic Review*, **67**, 297 308.

Eaton, B. C. and Lipsey, R. G. (1989). Product differen tiation. In R. Schmalensee and R. D. Willig (eds.), *Handbook of Industrial Organization*. New York: North-Holland.

Hotelling, H. (1929). Stability in competition. *Economic Journal*, **39**, 41 57.

Ireland, N. J. (1987). *Product Differentiation and Non Price Competition*. Cambridge, MA: Blackwell.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, **74**, 132 57.

Salop, S. C. (1979). Monopolistic competition with out side goods. *Bell Journal of Economics*, **10**, 141 56.

Schmalensee, R. (1978). Entry deterrence in the ready-to-eat breakfast cereal industry. *Bell Journal of Economics*, **9**, 305 27.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

product life cycle

Laura Power

From a variety of perspectives, researchers have hypothesized that products are characterized by patterns of evolution which have been called product life cycles. In most of these theories, new products evolve over time from "young" to "old," and the various stages of the cycle can be documented, because they are associated with certain observable characteristics (see Clark, 1985, for descriptions of the nature of the cycles; *see also* INDUSTRY LIFE CYCLE). Thus, life cycle theory highlights the importance of prod

uct, in addition to process, innovation. Product innovation implies that technology is dispersed through the introduction of new products, whereas process innovation implies that technology is introduced through new production processes. The continuum of product life cycles, arising from product innovation, potentially has implications for patterns of trade, STRATEGIC BEHAVIOR among firms, and even marketing strategies.

For example, Vernon (1966) suggests that international trade patterns are driven by the fact that new products are "born" in technologically superior developed countries, but when technology becomes standardized, production moves to less developed countries, where labor is cheaper. Based on the Schumpeterian notion of the innovative entrepreneur, other research investigates product life cycles in the context of strategic COMPETITION and innovation among firms. Recent marketing research simply tries to document the existence of product cycles using the implied observable characteristics of product cycle models. In practical terms, the existence of these product cycles can have implications for the speed at which manufacturers get new products on the market, the available recovery period for R&D, and the decision to innovate versus imitate. See Klepper (1992) for a full description and bibliography concerning product life cycles.

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Industrial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Clark, K. B. (1985). The interaction of design hierarchies and market concepts in technological evolution. *Research Policy*, **14**, 235.

Klepper, S. (1992). Entry, exit, and innovation over the product life cycle. Paper presented at the 1992 Conference of the International Joseph A. Schumpeter Society, Kyoto, Japan.

Segerstrom, P. S., Anant, T. C. A., and Dinopoulos, E. (1990). A Schumpeterian model of the product life cycle. *American Economic Review*, **80**, 1077.

Vernon, R. (1966). International investment and international trade in the product cycle. *Quarterly Journal of Economics*, **80**, 190.

product proliferation

Robert E. McAuliffe

When firms compete through PRODUCT DIFFERENTIATION, incumbent firms have incentives to offer many brands and "crowd the shelf." In 1972, the US ready to eat breakfast cereal manufacturers were charged with antitrust violations by the Federal Trade Commission. Among the charges was the claim that the cereal producers conspired to prevent ENTRY through product proliferation. The alleged strategy of the cereal producers was to introduce a large number of brands to fill the market in such a way that existing brands could earn ECONOMIC PROFIT and yet new entrants would not be able to enter the industry.

Schmalensee (1978) developed a theoretical model which showed how a product proliferation strategy might succeed. Scherer (1979) estimated the costs of introducing new cereal brands and compared those costs with the benefits to consumers of these brands to determine if the number of brands was excessive. He found that the benefits from the new brands were less than the costs and therefore the number of brands was not efficient (*see* OPTIMAL VARIETY).

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Industrial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Scherer, F. M. (1979). The welfare economics of product variety: An application to the ready-to-eat cereals industry. *Journal of Industrial Economics*, **28**, 113 34.

Schmalensee, R. (1978). Entry deterrence in the ready-to-eat breakfast cereal industry. *Bell Journal of Economics*, **9**, 305 27.

product variety

Kostas Axarloglou

Product variety is the number of variants in a family of products with similar characteristics, usually referred to as brands or models of the product. Product varieties can differ from each other in terms either of QUALITY or of other

characteristics unrelated to quality (*see* PROD
UCT DIFFERENTIATION).

As Eaton and Lipsey (1989) observed, there
are several *stylized facts* associated with product
variety:

- Many industries offer a large number of
  similar but differentiated products.
- The number of different varieties offered in
  the market is a small subset of the set of all
  possible varieties.
- In most industries, companies produce more
  than one variety of the same product.
- Each consumer purchases only a small subset
  of the available varieties in the market.
- Consumers appear to have different prefer
  ences, since they consume different varieties,
  and this cannot be explained solely by differ
  ences in their income.

The degree of product variety is determined by
the interaction of consumers, firms (cost of
production), and the MARKET STRUCTURE.
Economists model consumer choice either
through a representative consumer, who is as
sumed to have unlimited preference for different
brands of the same product, or through many
consumers, where each has a most preferred or
"ideal" brand to consume. Companies offer
product variety in order to increase their market
share and revenues, but their costs of production
also increase because of additional FIXED COSTS
associated with each brand and a more limited
market which prevents firms from fully realizing
ECONOMIES OF SCALE. Product varieties can
also be used strategically to deter ENTRY of
competitors and protect market shares (*see* BAR
RIERS TO ENTRY; PRODUCT PROLIFER
ATION). Finally, market structure influences
the degree of product variety in the market. In
oligopolies and monopolistically competitive
markets, companies introduce product varieties
to compete with other firms (*see* MONOPOLIS
TIC COMPETITION; OLIGOPOLY).

Overall, product variety will be greater

- the lower the fixed cost of introducing a
  variety in the market, since a new variety is
  more profitable the lower its fixed costs;

- the larger the market for the family of prod
  ucts in consideration, since large markets
  result in more revenues which can support
  the fixed costs of introducing more products;
- the weaker the degree of substitution be
  tween the different brands in the group,
  since the availability of SUBSTITUTES in
  creases the price ELASTICITY of market
  demand and lowers price markups (and
  profit margins; *see* MARKUP PRICING).

Product variety is also a significant component in
international trade. Countries with similar struc
tures may engage in trade in similar industries
(*intra industry trade*), since they produce and
trade different varieties in the same family of
products (see Kierzkowski, 1989). This is be
cause consumers demand product variety owing
to differences in tastes, but firms prefer to pro
duce few varieties (all else equal) when there are
economies of scale. Therefore, countries may
trade in different product varieties to satisfy con
sumers' preferences while allowing companies to
take advantage of economies of scale.

Product variety appears to have some impact
on the long run growth of economies (see Sto
key, 1988; Grossman and Helpman, 1991).
Finally, the amount of product variety in the
market may be influenced by the business
cycle, increasing during economic booms and
decreasing during economic recessions. Conse
quently, firms are willing to adjust their pricing
policies appropriately in response to these fluc
tuations in product variety over time (see Axar
loglou, 1993).

Bibliography

Axarloglou, K. (1993). Product variety, countercyclical
   price markups and short run macroeconomic fluctu-
   ations. Working paper, University of Michigan.
Eaton, B. C. and Lipsey, R. G. (1989). Product differen-
   tiation. In R. Schmalensee and R. D. Willig (eds.),
   Handbook of Industrial Organization. New York:
   North-Holland.
Grossman, G. M. and Helpman, E. (1991). *Innovation and
   Growth in the Global Economy*. Cambridge, MA: MIT
   Press.
Kierzkowski, H. (ed.) (1989). *Monopolistic Competition
   and International Trade*. New York: Oxford University
   Press.

Lancaster, K. (1990). The economics of product variety: A survey. *Marketing Science*, **9**, 189 206.

Stokey, N. L. (1988). Learning by doing and the introduction of new goods. *Journal of Political Economy*, **96**, 701 17.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

production functions

Robert E. McAuliffe

The production function for a firm shows the maximum output which can be produced with specific levels of inputs, given the available tech nology. If the firm's output is denoted by q, and the inputs in production are capital, K, labor, L, and other inputs, R, then the production func tion can be written as:

$$q = F(K,L,R)$$

where the function $F(.)$ relates the levels of inputs to the level of output. Production func tions have several properties (see Baumol, 1977). The first property is that the production func tion must exhibit DIMINISHING RETURNS. This means that if additional units of an input (such as labor) are added to production, holding the other inputs in production constant, eventu ally the increases in output will diminish (see LAW OF VARIABLE PROPORTIONS).

Another important property of production functions is returns to scale (or ECONOMIES OF SCALE). The issue here is the effect on output from increasing the amounts of all the inputs in production. For example, if all inputs in production are doubled, will output double or rise by a different amount? If doubling all inputs causes output to double, then there are CON STANT RETURNS TO SCALE. If doubling all inputs causes output to more than double, then there are increasing returns to scale. Finally, if doubling all inputs in production causes output to rise by less than double, there are decreasing returns to scale.

Production functions are also characterized by the degree of substitution between inputs in production. Some technologies require that inputs are used in fixed proportions, so that additional labor input cannot substitute for less capital input, for instance. The isoquant curves show the different combinations of inputs that are possible while keeping the output level con stant (*see* ISOQUANT ISOCOST CURVES). Normally, as the price rises for one input in production, firms will use less of that input and will substitute other inputs for the more expen sive input.

Bibliography

Baumol, W. J. (1977). *Economic Theory and Operations Analysis*. Englewood Cliffs, NJ: Prentice-Hall.

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

profit maximization

Gilbert Becker

Although the goal of profit maximization is not always consistent with that of sales (revenue) maximization and other goals that firms may have (*see* OBJECTIVE OF THE FIRM), profit maximization is generally identified as one of the primary goals of most businesses. Since the level of profit is measured as the TOTAL REV ENUE of the firm minus its TOTAL COSTS, profit maximization requires locating the level of output at which the difference between these two measures is a maximum. This level of output can be found by using marginal analysis where the MARGINAL REVENUE from one add itional unit sold is compared with the MAR GINAL COST of its production. The general rule for profit maximization requires that man agers find that level of output where marginal revenue and marginal cost are equal, since at this level of output no extra profits could be earned by increasing or decreasing production. At output levels where the marginal revenue from an additional unit to be sold exceeds the cost of its production, additional profits can be made by increasing the level of output. Similarly, if the marginal cost of the last unit already produced exceeded the additional revenue received from its sale, the marginal profit from that unit would be negative and its production and sale would

have reduced total profits. In this case, the firm's output level should be decreased as this would increase total profits.

This method is universally applicable to firms regardless of the basic market structure in which they compete (*see* BASIC MARKET STRUC TURES). Competitive firms, having no control over price, and monopolists, wherein manage ment supply decisions may significantly influ ence price, may both use this technique to assure the highest level of profits possible. In addition, many alternative applications of this method may be used. For example, managers of retail shops who are considering extending their daily store hours or opening for business on weekends need to examine the marginal revenues and mar ginal costs of these additional hours to assure that the profit maximizing number of hours are chosen. Similarly, with respect to hiring deci sions, choosing the optimal (profit maximizing) number of workers to hire requires the examin ation of the marginal gains and costs of add itional workers.

See also *economic profit; marginal product*

Bibliography

Maurice, C. and Thomas, C. (2002). *Managerial Econom ics*, 7th edn. Boston: McGraw-Hill Irwin.

Nicholson, W. (2000). *Intermediate Microeconomics and its Application*, 8th edn. New York: Harcourt.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

property rights

Laurence S. Moss

According to Armen Alchian, "a property right is a socially enforced right to select uses of an economic good." Lawyers think of property as "bundles" of separate property rights that can be bought and sold, leased, mortgaged, and par tially divided. In recent years the COASE THE OREM has encouraged economists to think about property more along the lines of the legal profes sion and thereby abandon or substantially modify older notions about production as com bining factors of production (labor, capital, and land) into output. All property rights

systems are valuable to decision makers when they are clearly defined and made secure with the assistance of infrastructures consisting of registrars, notaries, accountants, attorneys, sher iffs, and judges.

Property can take a myriad of forms in modern business situations, and combinations of property in the form of a business corporation can be represented by shares of common stock or can be held in a trust in the form of trust shares. New combinations and new legal forms are con ceivable as what Warren Samuels called the "legal economic nexus" undergoes change. From the business point of view, property func tions much as any typical business asset does (e.g., a machine) and its value is linked to the discounted PRESENT VALUE of its expected future flow of net benefits.

Well defined property combinations, espe cially when owned by the modern limited liability corporation, have permitted the accumulation of vast amounts of financial capital from individuals who otherwise would not find it acceptable to hold their wealth in this way. By these institu tional expedients, experimentation with new technologies and the creation of new business forms has resulted in the blossoming of new products and services that have become the hallmark of private property market systems.

The existence of property rights helps avoid disputes, especially about incompatible uses of economic goods, such as with air and water rights. Property rights can provide incentives to others to mitigate the situation that originally led to the conflict about planned uses. When property rights help mitigate disputes and align incentives, we speak of the "efficiency enhan cing" effects of property rights structures. For example, consider a small restaurant serving "fresh" fish to the public. The restaurant may impose side effects (negative EXTERNALITIES) on its neighbors such as odors. Furthermore, the neighbors may protest and demand that the local licensing authorities shut it down. If the author ities were to make it clear to the neighbors that the restaurateur does indeed have an inviolable "property right" to sell cooked fish dinners at this location, then the neighbors are free to ne gotiate, perhaps to purchase that restaurant or, more creatively, give the owners a monthly cash stipend provided that he does not cook fish on

certain days and/or between certain hours. Over time the neighbors may find it more economical to subsidize a better venting system for the restaurant or provide air conditioners to their homes, thereby eliminating the odors altogether.

According to Ronald Coase, Harold Demsetz, and other members of the law and economics school, such transactions, once property rights have been clearly defined and unhesitatingly enforced, will help maximize the value of output and thereby make the markets more efficient (see EFFICIENCY). In such cases the existence of clearly defined property rights is a precondition for the emergence of markets and the market processes associated with their presence.

As the market systems evolve, new forms of property emerge, often with the assistance of government but sometimes without. Some prop erty rights consist mostly of some person's promise to perform some act or acts at a future date. Buying and selling other people's prom ises, which was once restricted primarily to the banking and financial sector, has now expanded to create novel forms of property and new areas of entrepreneurial venturing. And so businesses are sold but always subject to conditions estab lished by government agencies such as Federal, State, local officials, and the courts. The buyer will not tender the purchase price for, say, a restaurant, until the appropriate government agency has promised to issue a special license and that he also obtain the "air rights" from the adjacent neighbor's property. In recent years, managers in different nations can avoid the risks of credit market conditions worsening and currency values falling by exchanging "swap promises," and an army of financial intermedi aries has emerged to facilitate these trades. When one person's promise is made under cer tain conditions, then secondary markets may emerge in which these promises are salable, thereby creating new forms of property and offering entrepreneurs new venturing opportun ities. These opportunities include the creation of more valuable asset combinations and incentives to speculate about their future value.

Socialist economists raise poignant questions as to what should or should not be subject to one person's exclusive dominion and control. Human beings have a moral claim to be regarded as more than mere bundles of property rights.

Exploring such insights, David P. Ellerman objected to well entrenched factory systems in which workers sell their "labor services" to the owners of property for a predetermined price (wages and salaries) and where the owners of property contrive to keep the results of that labor activity along with any profits and/or cap ital gains. Ellerman suggested that it is more consonant with western standards of morality and the meaning of language itself to have labor *hire* property and sell the outputs for their own account. In this way the traditional "employer– employee relation would have to be replaced by a system where each is self employed and this is the parallel of political self determination in democracy." To the extent that such radical thinking leads to a modification of customs, norms, and finally the legal system itself, the market system will assume a new shape and evolve in new directions.

One practical result of Ellerman's thinking is that modern managers may in the future become the agents of teams of employees who will need to find ways to protect their personal assets against the downside risks of business failure and tort liability. Under Ellerman's model, busi ness organizations and factory establishments will become more in the nature of massive tool and equipment rental operations. The workers become the lessees and the factory owners re ceive fixed prenegotiated rental prices. Regard less of how basic business institutions evolve and are restructured in the next millennium, it is reasonable to suppose that entrepreneurial ven turing will continue to flourish even in a worker style, capital using economy because the search for new combinations of property and their pur chase and sale remain indelible features of any market system.

Bibliography

Alchian, A. A. (1987). Property rights. In J. Eatwell, M. Milgate, and P. Newman (eds.), *The New Palgrave: A Dictionary of Economics*. New York: Macmillan.

Demsetz, H. (1964). The exchange and enforcement of property rights. *Journal of Law and Economics*, **7**, 11 26.

Dorn, J. A. (1994). The collapse of communism and post-communist reform. In P. J. Boettke (ed.), *The Elgar Companion to Austrian Economics*. Aldershot: Elgar.

Ellerman, D. P. (1992). *Property and Contract in Econom
ics: The Case for Economic Democracy*. Cambridge, MA:
Blackwell.

Rosenberg, N. and Birdzell, L. E., Jr. (1986). *How the
West Grew Rich: The Economic Transformation of the
Industrial World*. New York: Basic Books.

Samuels, W. J. (1989). The legal–economic nexus. *George
Washington Law Review*, **57**, 1556 78.

public goods problem

Lidija Polutnik

Public goods have two properties that differen
tiate them from private goods. First, they are
nonrival in consumption: once the good is pro
vided, the additional resource cost of another
person consuming the good is zero. In other
words, these goods are used and not consumed.
Second, public goods are nonexclusive: the con
sumption of a good is nonexcludable when it is
either very costly or impossible to prevent others
from consuming the good. Some examples of
goods with this property are national defense,
exploration of outer space, a television program,
and street lighting. However, the definition of a
public good is not absolute; it depends on market
conditions, the state of technology, and legal
arrangements. For example, highway availability
during rush hour traffic becomes rival in con
sumption and each driver creates a negative ex
ternality (*see* EXTERNALITIES). It is often
possible to exclude consumers from consuming
some public goods, e.g., cable television.

Since public goods can be consumed by more
than one consumer at the same time, the market
demand for a public good is defined as the verti
cal summation of the DEMAND CURVES of every
consumer of a public good. Samuelson (1954)
derived formal conditions for the efficient pro
vision of public goods. If public goods are pro
vided by private companies, then their level of
production will be suboptimal and correspond
ingly the incentives for overuse are greater. A
free rider problem occurs since consumers
cannot be excluded from the consumption of a
public good, and they may try to avoid paying for
their consumption of this good. In addition, it is
difficult to determine the true market demand
for public goods. At least since Samuelson, it has
been known that financing schemes such as
those proposals where an individual's tax is set
equal to his marginal benefit provide perverse
incentives for consumers to misrepresent their
preferences. Consumers are reluctant to reveal
high demand for public goods because they fear
they will have to pay a higher tax. Schemes that
are immune to such misrepresentations (in cer
tain circumstances) have been developed in
recent years (Clarke, 1971; Groves and Loeb,
1975).

The public goods problem can also be im
proved upon by changing, assigning, or creating
PROPERTY RIGHTS. When property rights are
secure and tradable and TRANSACTIONS COSTS
are low, the COASE THEOREM suggests that
people will trade rights until the new pattern of
ownership is efficient. For example, if everyone
is allowed to fish in a particular area, then no one
has property rights to the fish until they are
caught. When property rights are unclear,
markets are inefficient, and in the case of fisher
ies, substantial overfishing occurs. Uncon
strained private fishing ignores the social costs
of a smaller fish population in the future and
fishermen make decisions with respect to their
private benefits and costs of fishing today. The
problem could be improved upon by giving ex
clusive fishing rights to one group in the fishery.
In that case, the group would take into account
the effect of current fishing on the future popu
lation of fish to avoid overfishing (also called the
tragedy of the commons). This problem also
arises in companies when individuals or depart
ments have free access to a "common" resource,
such as a secretarial pool or a copy machine. No
one has an incentive to avoid overloading the
pool (machine), so secretaries (and the machine)
become overworked and the quality of service
declines.

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Indus
trial Organization*, 3rd edn. Reading, MA: Addison-
Wesley.

Clarke, E. H. (1971). Multipart pricing of public goods.
Public Choice, **11**, 17 33.

Groves, T. and Loeb, M. (1975). Incentives and public
inputs. *Journal of Public Economics*, **4**, 211 26.

Samuelson, P. A. (1954). The pure theory of public ex
penditure. *Review of Economics and Statistics*, **36**, 387 9.

Q

quality

Robert E. McAuliffe

Variations in product quality introduce several issues concerning the performance of markets. When consumers cannot determine the level of quality of a good before purchase, their IMPER FECT INFORMATION can cause a "lemons problem" where the market breaks down. This problem arises because, without additional in formation, consumers expect a low level of qual ity and ultimately only low quality sellers will remain in the market (*see* LEMONS MARKET).

Quality improvement has been viewed in three different ways in the literature (Levhari and Peles, 1973; Beath and Katsoulacos, 1991). First, quality increases may increase the demand for the product (such as ADVERTISING expend itures and other promotional expenses). The second approach considers quality as a perfect substitute for quantity. In these models, con sumers value the total level of services provided by the good (such as the total hours of light pro vided by a light bulb) and consider one 2 year light bulb as a perfect substitute for two 1 year light bulbs. Finally, quality has been treated as an increase in the durability of a product.

SOCIALLY OPTIMAL QUALITY LEVELS

Will free markets provide the socially optimal level of quality? In the case of the single product MONOPOLY (where only one level of quality can be selected), the answer depends on how the marginal consumer in the market values quality. The socially optimal level of quality in a market is determined by consumers' willing ness to pay for a quality increase on average. Therefore, an increase in quality is efficient if consumers in the market, on average, value that increase enough to pay the cost of providing it.

However, for a profit maximizing monopolist the question is not whether on average con sumers are willing to pay for an increase in quality, but whether the marginal consumer is willing to make that payment. So the level of quality in the market may be too high or too low depending on the marginal consumer's willing ness to pay for quality (see Tirole, 1988).

When firms compete in a market and can choose different levels of quality (there is *vertical product differentiation*), firms have incentives to avoid price COMPETITION by choosing differ ent quality levels. Thus when consumers in the market all rank the high quality product above a low quality product, but differ in their willing ness to pay for high quality, firms will try to differentiate their products as much as possible (*see* PRODUCT DIFFERENTIATION). If one firm has positioned itself as the high quality producer, then the next entrant will produce a low quality product to avoid price competition, as long as consumer preferences are sufficiently diverse (see Tirole, 1988). This incentive tends to create too much product differentiation in the market.

When quality is a perfect substitute for quan tity, will the level of quality provided in the market vary under different market structures (*see* MARKET STRUCTURE)? Since a monopolist is protected from competition, it had been thought that the level of product quality would be lower under monopoly than under PERFECT COMPETITION (Schmalensee, 1970; Beath and Katsoulacos, 1991). But Swan (1970) found that a monopolist would provide the same level of quality (measured as durability) as perfectly competitive firms if there are CONSTANT RETURNS TO SCALE in production and when consumers are perfectly willing to substitute quality (durability) for quantity. Under these

conditions, market structure does not affect the level of quality, although the result does not hold under different cost conditions or consumer preferences (see Liebowitz, 1982).

OPTIMUM DURABILITY AND PLANNED OBSOLESCENCE

An interesting problem arises when a monopolist produces a durable good that lasts for several periods because once the product is sold, the monopolist faces competition in subsequent periods from its past production in the resale market. In addition, the monopolist is tempted to pursue intertemporal PRICE DISCRIMINATION by charging a high price to the early buyers who most want the product in the initial periods, then lowering the price in later periods to sell to consumers who value the product less. Since consumers expect the monopolist to decrease price in later periods (because that is its profit maximizing choice), they are less willing to pay a high price today. Given this, Coase (1972) conjectured that monopoly profits would decrease as the length of time between price adjustments decreased, and this limits monopoly power. To avoid this problem, a monopolist would prefer to lease a product rather than sell it, and when this is possible, the monopolist will choose the socially optimal level of durability (see Tirole, 1988). But when the monopolist cannot lease the product, it has incentives to plan obsolescence and decrease the durability of the good. This occurs because demand in the current period will be higher if fewer units of last period's production can be used this period (see Bulow, 1986; Tirole, 1988).

Bibliography

Beath, J. and Katsoulacos, Y. (1991). *The Economic Theory of Product Differentiation*. Cambridge: Cambridge University Press.

Bulow, J. (1986). An economic theory of planned obsolescence. *Quarterly Journal of Economics*, **51**, 729 48.

Carlton, D. W. and Perloff, J. M. (2000). *Modern Industrial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Coase, R. (1972). Durability and monopoly. *Journal of Law and Economics*, **15**, 143 9.

Levhari, D. and Peles, Y. (1973). Market structure, quality and durability. *Bell Journal of Economics*, **4**, 235 48.

Liebowitz, S. (1982). Durability, market structure and new-used goods models. *American Economic Review*, **72**, 816 24.

Schmalensee, R. (1970). Regulation and the durability of goods. *Bell Journal of Economics*, **1**, 54 64.

Swan, P. L. (1970). Durability of consumer goods. *American Economic Review*, **60**, 884 94.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

R

R²

Theofanis Tsoulouhas

R² denotes the coefficient of determination in LINEAR REGRESSION. The coefficient of deter
mination is a measure of the goodness of fit of a regression to the data. The coefficient shows the proportion of the sample variation of the dependent variable that is explained by the re
gression of the dependent variable on the inde
pendent variables. For instance, an R² of 0.87 implies that the regression explains 87 percent of the variation in the dependent variable. In the context of a multiple regression, i.e., a regression with more than one independent variable, the coefficient is often called the "coefficient of multiple determination." Algebraically, R² is the ratio of the regression sum of squares (de
noted by RSS) to the total sum of squares (de
noted by TSS), i.e.,

$$R^2 = \frac{RSS}{TSS}; \; RSS = \sum (\hat{Y}_i - \bar{Y})^2;$$
$$\text{and } TSS = \sum (Y_i - \bar{Y})^2$$

where RSS is the variation of the estimated values of the dependent variable, TSS is the sample VARIANCE of the dependent variable, and $\bar{Y}$ is the sample (arithmetic) mean of the dependent variable. Therefore, the variation in a variable is the sum of the squared deviations of the variable from its mean. Note that all summa
tions are over the observations $i = 1, 2, \ldots, T$, where T is the sample size. Because TSS = RSS + ESS, where

$$ESS = \sum_{i\,1}^{n} \hat{v}_i^2$$

denotes the error sum of squares (which is the sum of the squared residuals of the regression), an alternative formula for R² is

$$R^2 = 1 - \frac{ESS}{TSS}$$

R² is a unit free measure and, by construction, its value is between 0 and 1. The closer R² is to 1, the better the explanatory power of the regres
sion.

Although R² is a valuable measure of the goodness of fit, it requires careful use for a number of reasons. First, R² does not account for the "degrees of freedom" of the analysis, where the degrees of freedom are defined as the sample size used in the regression (denoted here by T) minus the number of independent vari
ables in the regression (denoted by K) and minus 1 for the regression intercept (if included). As a result:

1 R² increases, or at least it does not decrease, even if variables which conceptually should not affect the dependent variable are in
cluded in the regression;
2 when the sample size is not large enough compared to the number of independent variables, then R² is likely to be high even when the independent variables are not strongly related to the dependent variable.

In addition, a high R² does not by itself imply that the regression model has been correctly specified (in the sense that no irrelevant vari
ables have been included in the regression, and no relevant variables have been omitted). For instance, a regression with TIME SERIES DATA that exhibit significant upward growth may have a high R² even though some important

explanatory variables have been omitted from the regression. Third, the interpretation of R^2 becomes difficult when the regression has no intercept. The reason is that without an intercept, the R^2 may not lie within the 0 to 1 range.

One way to correct some of the problems with R^2 is to modify the coefficient by adjusting for the degrees of freedom. The "adjusted" or "corrected" coefficient of determination, typically denoted by a "bar" on top of R^2, does just that. The adjusted coefficient is given by

$$\bar{R}^2 = 1 - \frac{\text{ESS}/(T - K - 1)}{\text{TSS}/(T - 1)}$$

where $T - K - 1$ is the number of degrees of freedom. Note that R^2 is always larger than or equal to the adjusted R^2, and also that the adjusted R^2 may even take negative values.

Bibliography

Greene, W. H. (2002). *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.
Gujarati, D. N. (2002). *Basic Econometrics*, 4th edn. New York: McGraw-Hill Irwin.
Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.
Newbold, P. (1995). *Statistics for Business and Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.
Pindyck, R. S. and Rubinfeld, D. L. (1991). *Econometric Models and Economic Forecasts*, 3rd edn. New York: McGraw-Hill.

real income

Robert E. McAuliffe

A consumer's real income is what that consumer can buy and it reflects that consumer's purchasing power or command over resources. Since INFLATION erodes the purchasing power of any amount of income, it is necessary to adjust nominal income (or current dollar) figures for inflation to obtain real, inflation adjusted income figures to determine the consumer's purchasing power. If the CONSUMER PRICE INDEX (CPI) reasonably measures changes in the cost of living for consumers in a given market, then the CPI can be used to adjust consumer income figures. Since the demand for many products is affected by changes in real income, adjusting nominal income figures for inflation can be important for forecasts and pricing decisions (see Douglas, 1992).

When the price of a product changes, it affects demand through the INCOME EFFECT and the SUBSTITUTION EFFECT. The income effect reflects the change in the consumer's real income after the price change and will be stronger the greater the proportion of the consumer's income spent on the product. For example, a 10 percent price increase in the price of soda will have a negligible income effect, but a 10 percent increase in the price of housing will generate a significant income effect (see Shughart, Chappell, and Cottle, 1994; Keat and Young, 2002).

Bibliography

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.
Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.
Shughart, W. F., Chappell, W. F., and Cottle, R. L. (1994). *Modern Managerial Economics*. Cincinnati, OH: South-Western Publishing.

real prices

Robert E. McAuliffe

According to economic theory, the decisions made by consumers and firms are based on real or relative prices. The real price of a good is its cost in terms of other products: how much a consumer must sacrifice in other goods to obtain this product. To calculate real prices in dollars it is necessary to adjust for INFLATION using some deflator such as the CONSUMER PRICE INDEX (CPI) or the producer price index. Dividing the current price by the appropriate deflator provides a measure of the real cost of the good in constant purchasing power base year dollars. Such a calculation shows managers the real impact on consumers of a given change in the price of the product (see Douglas, 1992).

Bibliography

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Shughart, W. F., Chappell, W. F., and Cottle, R. L. (1994). *Modern Managerial Economics*. Cincinnati, OH: South-Western Publishing.

rent seeking

Kent A. Jones

Broadly defined, rent seeking is an attempt by the owner or potential owner of an asset to secure excess returns on the asset by investing resources in influencing public policy. Krueger (1974) developed this concept originally in terms of the efforts by firms to capture MONOPOLY profits by acquiring government issued licenses on quotas that grant exclusive access to a market. One can extend this idea to the owner of a factor of production whose price is affected by government intervention in the market for the factor itself or for the final product in which the factor is an input. Bhagwati (1982) developed the broader definition of rent seeking to include what he termed "directly unproductive profit seeking (DUP) activities." To the extent that public policy can affect prices for assets or final products, the rent seeker has a strong incentive to invest resources to influence policy decisions in an attempt to alter market driven prices or access to the market in his favor.

The problem of rent seeking stems from the general asymmetry in welfare effects of public policy: the benefits of government market intervention tend to be concentrated on small groups of identifiable beneficiaries, while its costs tend to be spread more thinly across the population. Trade policy provides the most salient examples of rent seeking in modern economies. Since a tariff or import quota can reduce import COMPETITION and allow domestic producers to raise prices and increase the returns on CAPITAL, labor, or other factors of production, for example, firms and workers in a domestic import competing industry are often motivated to lobby in favor of the trade restriction. In contrast, consumers, whose collective losses from the trade restriction are greater than the gains of the protected industry, are much more difficult to organize for the purposes of a counter lobby.

If the government restricts trade through a quota arrangement, further opportunities for rent seeking arise as individuals invest resources to try to secure control over the quota rights, which give the holder the ability to raise the price of the imported product and thereby capture the scarcity premium associated with the trade restriction. Other examples include efforts by firms to acquire exclusive government issued monopoly rights over an existing product and the expansion of productive capacity by firms in order to qualify for increased government allocations of inputs. In general, most government regulation of markets tends to create rent seeking activity in the broader sense.

Rent seeking has social costs, since it tends to dissipate above market returns in the process of competition to get them (Friedman, 1990: 476). Firms and individuals who can benefit from changes in government policies are often willing to divert real resources toward influencing policy makers. The OPPORTUNITY COSTS of such activities include not only the explicit financial outlays associated with lobbying (which presumably have their next best alternative use in market driven productive activities by the individual or firm), but also the implicit value of time and effort diverted from the "normal" activities of market participants toward the acquisition of political influence. Theoretically, the entire amount of excess returns associated with the government policy could be dissipated by rent seeking activity. This consideration suggests that the social cost of government regulation should include a measure of the wasted resources due to rent seeking.

See also *deadweight loss; monopoly*

Bibliography

Bhagwati, J. N. (1982). Directly unproductive profit-seeking (DUP) activities. *Journal of Political Economy*, **90**, 988 1002.

Friedman, D. (1990). *Price Theory*, 2nd edn. Cincinnati, OH: South-Western Publishing.

Krueger, A. O. (1974). The political economy of the rent-seeking society. *Journal of Political Economy*, **64**, 291 303.

rents

Lidija Polutnik

Economic rents are defined as payments to entre preneurs and owners of resources which are over and above the resource's opportunity cost (*see* OPPORTUNITY COSTS). For example, an eco nomic rent is a portion of payment in excess of the minimum amount of profit necessary for a firm to enter a particular industry or a worker to accept a particular job. These excess payments are called economic rents because of the analogy to land rents: a given piece of land is fixed in supply, and whatever rent is received for it is by its nature a surplus over what is needed to keep the land in use. Often economic rents exist be cause of legal restrictions on ENTRY into the industry. For example, the number of taxicab licenses and liquor licenses is often fixed by law, thus restricting entry to the industry and creating MONOPOLY profits (economic rents). Occupational licensing laws which restrict labor supply and thereby increase the market wage also confer economic rents to licensees. In addition, economic profits which may appear to exist in perfectly competitive markets in the long run may actually be economic rents (*see* ECONOMIC PROFIT; PERFECT COMPETITION).

The existence of economic rents may lead to profit seeking and to rent seeking behavior (Brickley, Smith, and Zimmerman, 2000; *see* RENT SEEKING). Entrepreneurs and owners of resources are selfishly motivated to reallocate their resources and find opportunities to put themselves in positions in which they would earn economic rents. Profit seekers are induced to enter the particular industry as a result of the existence of profits as well as to reallocate their scarce productive resources to their most highly valued uses. As a result of these decisions, output expands, prices decline, and in the long run eco nomic profits are competed away and society's welfare is improved. However, rent seeking refers to the firms who tend to spend money and exert effort to acquire and maintain the mon opoly position or claims to resources that are in fixed supplies. In this case resources may be used in an unproductive way from the society's point of view and its welfare may worsen.

Quasi rents are defined as SHORT RUN economic rents earned by the firm's fixed productive resources. Perfectly competitive firms in the short run are willing to produce as long as price is larger than AVERAGE VARIABLE COSTS. The difference between the price and average variable cost is called quasi rent because it is similar to an economic rent in the short run, but not in the long run. Whereas economic rents are defined in terms of decisions to enter a job or an industry, quasi rents are defined in terms of a decision to EXIT (Milgrom and Roberts, 1992). These quasi rents are commonly created by costs associated with changing businesses, jobs, or careers, such as the cost of moving or specific on the job training. These costs cannot be re covered by the worker who chooses to change jobs because specialized human capital has been developed which cannot be used in another com pany. Quasi rents represent a normal return on past investments and can continue to exist in long term relationships given a proper contract design (*see* CONTRACTS).

Bibliography

Brickley, J. A., Smith, C. W., and Zimmerman, J. L. (2000). *Managerial Economics and Organizational Archi tecture*. Chicago: McGraw-Hill.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Milgrom, P. and Roberts, J. (1992). *Economics, Organiza tion and Management*. Englewood Cliffs, NJ: Prentice-Hall.

reservation price

Robert E. McAuliffe

A consumer's reservation price is the highest price that the consumer would be willing to pay to purchase one unit of a product rather than forgo the purchase. On the supply side of the market, the reservation price would reflect the minimum acceptable price at which a supplier would be willing to sell the product or service. The difference between the price a consumer pays for a product and the reservation price is the CONSUMER SURPLUS for that consumer, while the difference between the price received by the supplier and its reservation price is the PRODUCER SURPLUS. Normally demand

functions measure the maximum quantity con sumers will purchase at a specific price (*see* DEMAND FUNCTION), but as Friedman (1976) has observed, this interpretation depends on the alternatives available to consumers. For example, conventional DEMAND CURVES assume con sumers are able to buy the specified quantity or less at the given price. However, the demand curve and the corresponding consumers' reserva tion prices will be different (and more likely higher) if the choice is to purchase the specified amount or none. Similarly, the reservation price for suppliers will be different (and more likely lower) if the purchase offer is an all or nothing bid.

Bibliography

Friedman, M. (1976). *Price Theory*, 2nd edn. Chicago: Aldine.

Jehle, G. A. and Reny, P. J. (2000). *Advanced Microeco nomic Theory*, 2nd edn. Boston: Pearson Addison-Wesley.

Varian, H. R. (1992). *Microeconomic Analysis*, 3rd edn. New York: Norton.

residual demand

Robert E. McAuliffe

The residual demand curve is the individual firm's demand curve which is that portion of market demand that is not supplied by other firms in the market. Thus it is the market DEMAND FUNCTION minus the quantity sup plied by other firms at each price. If market demand is $D(p)$ and the supply of other firms is $S_o(p)$, then the residual demand for firm i is:

$$D_i(p) = D(p) - S_o(p) \qquad (1)$$

The relationship between residual demand and market demand can also be expressed in terms of elasticities (*see* ELASTICITY). For example, if there are n firms in an industry which are all identical, then the elasticity of residual demand facing firm i can be expressed as:

$$\varepsilon_i = \varepsilon_p \times n - \varepsilon_{so} \times (n-1) \qquad (2)$$

where ε_p is the price elasticity of market demand, and ε_{so} is the elasticity of supply of the other firms in the market, which will be positive (see Carlton and Perloff, 2000). As one might expect, the elasticity of residual demand facing a single firm is much higher than the elasticity of demand for the market, so a single firm will find its demand very sensitive to price changes. In addition, an individual firm's elasticity of demand will be higher (in absolute value) the greater the elasticity of supply of rival firms in the market, since there will be more substitutes available to firm i's customers when this supply elasticity is high.

When only one firm serves the market ($n = 1$), the firm is a MONOPOLY and the firm's demand is the same as market demand. For a DOMINANT FIRM facing a number of small price taking competitors, the elasticity of residual demand is

$$\varepsilon_d = \frac{Q}{Q_d} \times \varepsilon_p - \frac{Q_o}{Q_d} \times \varepsilon_{so} \qquad (3)$$

where ε_d is the elasticity of residual demand for the dominant firm, Q_d is the output of the dominant firm, Q is industry output (which is the sum of the output produced by the dominant firm and its rivals), and Q_o is the output of the other smaller firms in the market. Here, the MARKET POWER of the dominant firm falls the higher the elasticity of supply of the other firms in the market, and this forces the dominant firm to choose a lower price. The dominant firm will also have less market power the smaller the proportion of industry output it produces.

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Indus trial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Stigler, G. J. (1968). *The Organization of Industry*. Chicago: University of Chicago Press.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

risk aversion

Roger Tutterow

Decisions involving risk are modeled as lotteries consisting of a set of possible outcomes and a probability distribution across these outcomes.

In this environment, optimal decisions depend not only upon the outcomes and their associated probabilities, but also upon a decision maker's attitude toward risk. An agent is said to be risk averse if he prefers the EXPECTED VALUE of a lottery to the lottery itself.

For example, consider a simple lottery of flipping a coin for which one receives $20 if the coin turns up "heads" and $0 if the coin is a "tails." The mathematical expectation or expected value of this lottery is $10. Facing a choice between the coin flip or accepting $10 with perfect certainty, a risk averse agent will strictly prefer the certain $10. In fact, a risk averse agent will accept some amount less than $10 rather than accepting the coin flip. The amount with which the agent is indifferent to the lottery is the CERTAINTY EQUIVALENT of the lottery.

In contrast, an agent who strictly prefers the coin flip to the certain $10 is said to be risk preferring. A decision maker who is indifferent between the two prospects is said to demonstrate RISK NEUTRALITY.

The primary framework for the analysis of choices involving risk is the expected utility hypothesis associated with von Neumann and Morgenstern (1944). In their theory, if an agent meets certain rationality axioms, then his preferences over outcomes may be characterized by a continuous utility function. Then, an agent will evaluate lotteries in terms of expected utility rather than expected value. If an agent is risk averse, his utility function will be concave in income. In contrast, risk preferring and risk neutral agents have utility functions which are convex and linear, respectively.

Since attitude toward risk is subjective, it is not surprising that agents may differ in the degree of risk aversion. Since risk averse agents have concave utility functions, one might expect the curvature of the utility function to relate to the degree of risk aversion. Following Arrow (1971) and Pratt (1964), risk aversion may be measured by the following coefficient of absolute risk aversion:

$$\text{ARA} = -u''/u' \qquad (1)$$

where u'' is the second derivative of the utility function with respect to income and u' is the first derivative. This coefficient is a measure of the concavity of the utility function that is scaled to be invariant to linear transformation of the utility function. Friedman and Savage (1948) argue that individuals may be risk averse over some ranges of wealth while risk preferring over other ranges. Empirical evidence suggests that most individuals become less risk averse as their income increases.

Risk aversion has a crucial role in the theory of both insurance and financial securities markets. Insurance markets exist so that risk averse agents may spread their risk among market participants. That insurance purchasers are willing to pay premiums in excess of the expected value of their losses demonstrates their risk aversion. In financial markets, investors require a higher return from risky securities. The latter example shows that risk aversion does not imply that an individual will not accept risky propositions, but rather that he or she will require additional compensation to do so.

Bibliography

Arrow, K. J. (1971). *Essays in the Theory of Risk Bearing*. Amsterdam: North-Holland.

Friedman, M. and Savage, L. J. (1948). The utility analysis of choices involving risk. *Journal of Political Economy*, **56**, 279 304.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Pratt, J. (1964). Risk aversion in the small and in the large. *Econometrica*, **32**, 122 36.

von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

risk neutrality

Roger Tutterow

In choices involving risk, optimal decisions depend upon the set of possible outcomes, the probability of their occurrence, and the decision maker's attitude toward risk. A decision maker is said to be "risk neutral" if he evaluates decisions in terms of the EXPECTED VALUE of the outcomes. For a risk neutral agent, the expected value and CERTAINTY EQUIVALENT of a lottery are equal.

There is little reason to expect that an indi vidual should display risk neutrality. However, Arrow and Lind (1970) showed that if returns on an investment are spread over a large population and are distributed independently of any aggre gate risk, then one may evaluate the investment as if risk neutral. A variant of this risk spreading argument is frequently used to defend risk neu trality in firms owned by a large number of STOCKHOLDERS.

Bibliography

Arrow, K. and Lind, R. (1970). Uncertainty and the evaluation of public investment decisions. *American Economic Review*, **60**, 364 78.

Varian, H. R. (1992). *Microeconomic Analysis*, 3rd edn. New York: Norton.

risk premium

Roger Tutterow

In the theory of choice under risk, a risk premium is defined as the difference between the EXPECTED VALUE and the CERTAINTY EQUIVALENT of a lottery. One may consider the risk premium as the compensation necessary to entice a risk averse agent to accept a risky proposition. Pratt (1964) demonstrated that for "small" gambles, the risk premium is proportion ate to the VARIANCE of the outcomes, the pro portion varying with the decision maker's degree of RISK AVERSION. Further, it can be shown that the more risk averse the agent, the larger the risk premium associated with a given lottery.

In financial markets, risk premiums exist as the additional return necessary to entice invest ors to accept risky investments. In debt markets, yields include premiums for default and inter est rate risk. The returns on equity securities include a premium for uncertainty about re sidual cash flows (*see* CASH FLOW).

Bibliography

Pratt, J. (1964). Risk aversion in the small and in the large. *Econometrica*, **32**, 122 36.

Varian, H. R. (1992). *Microeconomic Analysis*, 3rd edn. New York: Norton.

S

search costs

Robert E. McAuliffe

These are the costs incurred by consumers as they try to find those products which best satisfy their needs. Search costs include the time spent looking for a product, finding the best price, and the costs of purchasing brands which failed to meet the consumer's expectations. These costs will be higher the greater consumers' REAL INCOME (because the opportunity cost of lost time is higher; *see* OPPORTUNITY COSTS), the more dispersed prices are in the market (because more effort would be required to find the best price), the more geographically dispersed sellers are in the market, and the less information consumers have initially.

Since search costs may reduce sales, institutions exist in markets to reduce those costs to potential buyers. Shopping malls reduce search costs by providing a central location for consumers to find sellers, firms may reduce search costs through ADVERTISING expenditures, establishing reputations for QUALITY or providing free samples to consumers. Consumers also receive information from independent sources such as Consumer Reports and Underwriters' Laboratories that help to overcome IMPERFECT INFORMATION in the market.

See also *search goods; transactions costs*

Bibliography

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.
Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.
Shughart, W. F., Chappell, W. F., and Cottle, R. L. (1994). *Modern Managerial Economics*. Cincinnati, OH: South-Western Publishing.

search goods

Robert E. McAuliffe

Search goods are those products whose QUALITY can be determined by inspection before purchase. The distinction between search goods and EXPERIENCE GOODS was developed by Nelson (1970, 1974), who suggested that ADVERTISING played an important role in providing information to consumers in markets where information may be imperfect (*see* IMPERFECT INFORMATION). Since consumers can verify the truthfulness of advertising claims for search goods before purchasing the product, Nelson suggested that the information provided by advertising for search goods was more likely to be truthful. In addition, the content of advertising messages for search goods was likely to be more factual and direct than for experience goods, and this implied that search goods would be more heavily advertised in the print media (newspapers and magazines) rather than in broadcast media, and that local media outlets would be used in greater proportion than national media outlets since local media would provide more directly relevant information to consumers.

Nelson found that his classification of products into search and experience categories allowed him to explain differences in advertising–sales ratios between products and in the media used by different products. Ehrlich and Fisher (1982) extended the model of advertising

in the market for information and weighed the costs and benefits of different methods of pro viding information to the market. For example, the potential buyers of producer goods tend to be small in number, very knowledgeable, and bene fit more from information provided through trade shows and salespeople. But the potential buyers of consumer products are more numer ous, less knowledgeable, and more heteroge neous, so broadcast media may be the most cost effective means to provide information to this audience. Since information can be effi ciently supplied through a variety of channels in the producer goods industries, Ehrlich and Fisher also predicted that advertising for these products would be less intensive than for con sumer goods.

Bibliography

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Ehrlich, I. and Fisher, L. (1982). The derived demand for advertising: A theoretical and empirical investigation. *American Economic Review*, **72**, 366 88.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, **78**, 311 29.

Nelson, P. (1974). Advertising as information. *Journal of Political Economy*, **82**, 729 54.

Sherman Act

Gilbert Becker

The Sherman Act of 1890 is the first major US antitrust law (*see* CLAYTON ACT; FEDERAL TRADE COMMISSION ACT). Written in an era of increasing industrialization, the Sherman Act was in part an effort to quell the attempts to increase MARKET POWER and the conse quences arising therefrom. Section 1 of the Act proscribes "Every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce" (see Stelzer, 1986: 594–8 for a more complete statement of this Statute). Although debate continues as to the meaning of this phrase, managers should be aware of certain definitive interpretations which the courts have consistently maintained. The first is that horizontal price fixing, i.e., any agreement between rival firms to coordinate prices, is in violation of the Act. Beginning with *US* v. *Trenton Potteries Company* in 1927, the court has strictly held the position that price fixing is *per se* illegal, meaning that the behavior is in and of itself a violation of the law. Thus, defenses of the fairness of the fixed prices, the need for price fixing during an economically unstable period for the industry, and other argu ments have consistently been struck down. The court has insisted that market forces shall deter mine what constitutes reasonable prices.

Evidence of illegal price fixing includes sharing price lists, discussing prices through trade associations, and more recently SIGNAL ING price information through computer net works, as was alleged to have occurred in the airline industry. Controversy still exists con cerning pricing in OLIGOPOLY markets where differing forms of PRICE LEADERSHIP are common and where conscious awareness of par allel pricing behavior among rivals is often routine. The difficulty herein lies with the recognition that quite often a firm has no choice but to follow a parallel pricing strategy – in reaction, for example, to a rival's price decrease – in order to avoid losses. Thus, what may appear to be a conspiracy is not necessarily one. To date the courts have typically required hard evidence beyond parallel pricing in order to infer a price fixing conspiracy.

Resale price maintenance, sometimes known as vertical price fixing, is an entirely different pricing agreement whereby a manufacturer es tablishes minimum prices at which retailers can resell its products. Prior to 1975 this type of price fixing was exempt from the Sherman Act. Allowing the institution of a minimum resale price was seen as a method to protect small retailers from the growth of large discount outlets which would otherwise have been able to offer lower prices than their smaller rivals. With the repeal of the laws protecting resale price maintenance, vertical price fixing currently is illegal, perhaps to the benefit of consumers who are able to find lower prices.

Enforcement of the Sherman Act concerning vertical price fixing has varied in the past two decades. During the 1980s the federal govern ment showed little interest in initiating cases against vertical price fixers. In 1995 though, a case was brought against Reebok International

Corporation, concerning the resale of its athletic footwear. Reebok agreed in a consent decree to not impose required minimum prices on retail ers, nor to threaten retailers with termination of the supply of its footwear in instances where the retailer does not conform to the manufacturer's suggested minimum resale price. Thus, man agers should be aware that a manufacturer's policy of identifying a resale price is not illegal as long as it is merely suggested and not imposed upon the retailer as a condition of the contractual agreement.

In addition to horizontal price fixing, market sharing agreements among rivals have also been proscribed by the courts in their interpretation of Section 1 of the Sherman Act. These arrange ments, whereby would be rivals agree to serve only certain customers, or avoid each other's geographic territories, are also *per se* illegal. This type of activity is often a significant com ponent of the behavior of CARTELS, which are also illegal under US antitrust law.

Section 2 of the Sherman Act indicates that "Every person who shall monopolize, or attempt to monopolize, or combine or conspire . . . to monopolize any part of the trade or commerce" is in violation of the law. Here again managers should be aware of several important and long standing interpretations of this Section by the US courts. First, the law identifies the act of *monopolizing*, rather than the mere existence of MONOPOLY, as being proscribed. Beginning perhaps with the 1914 case of *US* v. *Standard Oil Company of New Jersey*, the Supreme Court has recognized and maintained this distinction. In that case, the court found that the firm had intended by its actions to dominate the market by driving out existing competitors from the industry. By contrast, the court also indicated in that case that the achievement of a monopoly position through pro competitive business prac tices such as the production of a superior product would not be a violation of the law. Moreover, monopolizing also includes anticom petitive acts designed to *maintain* a monopoly position once it has been achieved, in part by excluding other potential competitors. This is a central finding in the recent MICROSOFT ANTITRUST CASE.

Second, the courts have indicated that having a pure monopoly position (i.e., a 100 percent

market share) is not a necessary requirement for finding a violation. The trusts in the major early cases involving the oil and tobacco indus tries were found to have approximately 90 per cent market shares. In the 1945 case of *US* v. *ALCOA*, the firm was found to be in violation of Section 2 even though, by some measures, it held only 60 percent of the market. By contrast, the US Steel Corporation in 1920 held slightly more than 50 percent of the national market, but was considered by the courts not to have achieved sufficient market power, and in part for this reason was found not to be in violation of the law.

A third important interpretation also evolved from the US Steel case. The court made clear the distinction between the absolute size of a firm and its size relative to the market. It held that absolute size does not necessarily convey market power and thus is not by itself an offense.

As a result, the courts appear to have created a two part test for establishing a violation of Section 2 of the Act. First, prima facie evidence of the existence of market power through holding a significant share of the relevant market (seemingly in excess of 60 percent) must be present. Second, evidence of anticom petitive attempts to monopolize must be demon strated.

The Sherman Act has recently fallen under sharp criticism concerning its ability to generate appropriate results in the "new economy" of the twenty first century. Critics, especially those defending Microsoft Corporation in its current antitrust lawsuits, argue that new economic con cepts such as NETWORK EXTERNALITIES pre sent challenges to the logic and appropriateness of the two part test which has long been a stand ard in Section 2 cases.

See also *antitrust policy (US); antitrust remedies (US); Microsoft antitrust case: remedies*

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Indus trial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Jorde, T., Lemley, M., and Mnookin, R. (1996). *Gilbert Law Summaries: Antitrust*, 9th edn. Chicago: Harcourt Brace Legal and Professional Publications.

Stelzer, I. (1986). *Selected Antitrust Cases: Landmark Decisions*, 7th edn. Homewood: Irwin.

Waldman, D. and Jensen, E. (2001). *Industrial Organization: Theory and Practice*, 2nd edn. Boston: Addison-Wesley.

short run

Robert E. McAuliffe

The short run is defined as that period of time during which at least one input in production is fixed and cannot be changed. Frequently CAPITAL, such as plant and equipment, is treated as fixed in the short run while labor and materials are considered to be variable, but long term labor CONTRACTS or unions may make some labor costs fixed. Given that at least one input is fixed in the short run, DIMINISHING RETURNS will apply to production.

As Stigler (1966) and De Alessi (1967) have noted, the language distinguishing the short run from the long run is ambiguous. For very short periods of time, all factors of production are fixed. For slightly longer periods of time, all factors of production are variable if no expense were spared. De Alessi argues that firms will adjust different inputs in the "short run" depending on the relative costs and benefits of a particular change given the production technologies available. The costs of adjusting factors of production should be higher the shorter the time interval under consideration, and for some inputs those costs will increase much more rapidly than for others. Therefore it may be more appropriate to think of the short run as the period of time when the firm chooses not to adjust some factors of production because it would significantly raise the costs of those factors relative to the expected benefits.

Bibliography

De Alessi, L. (1967). The short run revisited. *American Economic Review*, **57**, 460 1.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Stigler, G. J. (1966). *The Theory of Price*, 3rd edn. New York: Macmillan.

short run cost curves

Robert E. McAuliffe

The SHORT RUN is defined as that period of time during which at least one input in production is fixed and cannot be changed. Frequently CAPITAL, such as plant and equipment, is treated as fixed in the short run while labor and materials are considered to be variable, but long term labor CONTRACTS or unions may make some labor costs fixed as well. Given that at least one input is fixed in the short run, DIMINISHING RETURNS will apply to production.

Following convention and to simplify the discussion, consider a firm using only two inputs in production where capital will be treated as the fixed factor of production and labor will be considered variable. This means that in the short run, the amount of capital the firm has available cannot be altered from its fixed level K_0 and the short run production function (*see* PRODUCTION FUNCTIONS) for the firm is then:

$$q = F(L, K_0) \qquad (1)$$

where q is the quantity of output produced and L is the quantity of labor employed. Total production costs for the firm will be equal to the wage rate (w) times the amount of labor employed (assuming labor can be hired at a constant wage) plus the fixed amount of capital times the economic cost of capital (r). Total costs are then:

$$TC = w \times L + r \times K_0 \qquad (2)$$

The first term on the right hand side of equation (2) is the total labor cost for the firm and these are the firm's total variable costs in this example (*see* TOTAL VARIABLE COST). The second term on the right is the firm's economic cost of capital, which is a fixed cost and does not vary with output (*see* FIXED COSTS). The firm's short run cost curves can then be derived by observing the changes in total costs (or the changes in total variable costs) as output changes. Since capital inputs are fixed in the short run, the firm must adjust labor inputs to change output, and the output which will be produced is determined by the production function in equation (1) above. In this example, total variable costs are just the labor costs in equation (2) above, so

marginal costs (*see* MARGINAL COST) are then equal to

$$MC = \frac{\Delta TVC}{\Delta q} = \frac{w \times \Delta L}{\Delta q} = w \times \left\{ \frac{\Delta L}{\Delta q} \right\}$$
$$= \frac{w}{MP_L} \quad (3)$$

where ΔTVC is the change in total variable costs, Δq is the change in output, and the term in braces on the right is the inverse of the MARGINAL PRODUCT of labor (MP_L). The marginal cost curve is the slope of the total variable cost curve at different output levels and since the wage rate is assumed to be constant, marginal costs will vary with changes in the marginal product of labor. Variable and marginal costs will be zero if no labor is hired and no output is produced. If one worker were hired, the variable and marginal costs of producing the first few units of output would be high (because the marginal product of a single worker would be low), but as additional workers are hired, the marginal product of labor increases, which causes marginal costs to decrease. Eventually, the marginal product of labor will reach a maximum value which, in this example, corresponds to the minimum value of the marginal cost curve. After this (point A in figure 1), the marginal product of labor decreases and the marginal cost of production increases.

The corresponding short run average cost curves can also be derived from the underlying production function and prices of inputs (equations (1) and (2) above). There are three short

run average cost curves: average total costs (SRATC), average variable costs (SRAVC), and average fixed costs (SRAFC). These average cost curves are given by:

$$SRATC = \frac{TC}{q} = \frac{(w \times L + r \times K_0)}{q} \quad (4)$$

$$SRAVC = \frac{TVC}{q} = \frac{w \times L}{q} \quad (5)$$

$$SRAFC = \frac{TFC}{q} = \frac{r \times K_0}{q} \quad (6)$$

where TC is the firm's total costs in equation (1) above (including the opportunity cost of capital; *see* OPPORTUNITY COSTS), TVC the firm's total variable costs, and TFC the total fixed costs of the firm. Of these average cost curves, average fixed cost is the least important in economics because fixed costs do not affect short run decisions. The short run marginal, average, and average variable cost curves are depicted in figure 1.

The marginal cost curve intersects both average cost curves at their respective minimum points (B and C), and as marginal costs continue to increase, the average cost curves rise but at a slower rate. As long as the marginal cost curve lies below the average variable and average total cost curves, the average cost curves must be falling because the cost of producing the last unit (the marginal cost) is less than the average cost of all the preceding units produced, and this pulls the average cost curves down. Similarly,
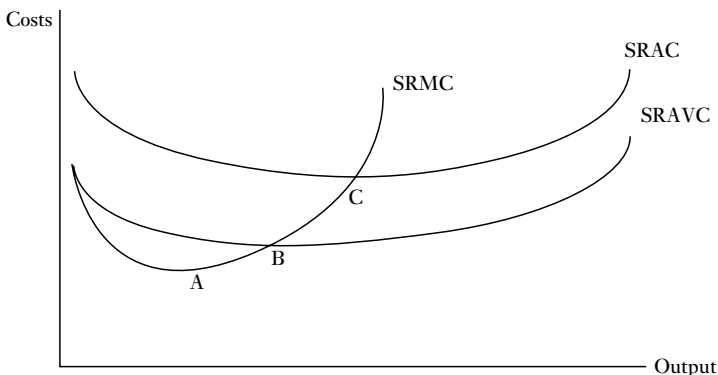
Figure 1 Short run cost curves

when the marginal cost curve lies above the average variable and average total cost curves, it pulls them up since the cost of the last unit exceeds the average of the previous units pro duced and adding any number to an average which exceeds the average must force the aver age value to rise.

The average variable cost curve plays a role in short run decisions because the firm does not have to incur these costs and can choose to produce no output. Therefore, if the price the firm will receive for its output is too low, the revenues earned will not cover the firm's vari able costs and the firm should shut down and produce nothing. Since fixed costs must be paid whether or not the firm produces output, these costs should have no effect on a manager's deci sion to produce in the current period. But the firm must pay variable costs only when output is produced, and if those variable costs cannot be paid from the revenues generated from produc tion, there is no reason to incur them. If the revenues can cover variable costs, then any add itional earnings can be applied to fixed costs and reduce the firm's short run losses. A perfectly competitive firm will maximize profits (*see* PER FECT COMPETITION; PROFIT MAXIMIZA TION) by producing output where price equals marginal cost for those points where marginal cost lies above the short run average variable cost curve (the marginal cost curve above point B in figure 1).

Theoretically, the average total cost curve, ATC, serves as a benchmark to predict industry trends. As mentioned above, the ATC curve includes the opportunity cost of capital, which is the normal, risk adjusted rate of return that could be earned if the firm's capital were employed in its next best use. When the firm's TOTAL REVENUE exceeds its TOTAL COSTS (which include the normal return on capital), the firm is earning an ECONOMIC PROFIT. This occurs when the price of the product (which is average revenue) lies above the ATC curve at the firm's profit maximizing output level. Therefore, when the product's price lies above the ATC curve, firms can earn higher returns in this industry than elsewhere and, if there are no BARRIERS TO ENTRY, ENTRY should occur in the long run. Similarly, if the price is below the firm's ATC curve, firms in

this industry are earning less than they could earn elsewhere and there should be EXIT from this industry in the long run.

Unfortunately, as Stigler (1966) and Fried man (1976) have observed, average total costs (which include the opportunity cost of capital) will change as industry demand changes. For example, an increased demand will cause profits to rise and increases the value of the firm's assets. Or a firm might have access to better resources or a better location which lower its costs of production and allow the firm to enjoy a return apparently above the normal competi tive rate. However, in both cases these resources could be sold for a capital gain and this is an opportunity cost to the firm. Therefore, if the firm valued its assets at this higher (market) value, its true long run average costs would be higher and its rate of return would be lower (*see* RENTS). What appear to be above normal profits may occur because of the failure to prop erly capitalize the value of the firm's assets.

Bibliography

Friedman, M. (1976). *Price Theory*, 2nd edn. Chicago: Aldine.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Stigler, G. J. (1966). *The Theory of Price*, 3rd edn. New York: Macmillan.

Shughart, W. F., Chappell, W. F., and Cottle, R. L. (1994). *Modern Managerial Economics*. Cincinnati, OH: South-Western Publishing.

signaling

Lidija Polutnik

When information is not readily available indi viduals can take the initiative to produce signals that reveal their attributes. Spence (1974) pro vided a systematic explanation of investment in education as a means of signaling productivity to the market. Suppose there are two types of workers: high ability workers whose marginal product of labor is 2 and low ability workers whose marginal product of labor is 1. Workers may invest in education or a degree at positive cost but, by assumption, education does not

increase on the job productivity. If information about each individual's productivity were avail able to the employer, workers would be paid their respective MARGINAL PRODUCT. When there is ASYMMETRIC INFORMATION, where em ployers do not know their prospective workers' marginal products and workers do, the employer pays everyone the same wage. That wage would be equal to the expected productivity of a ran domly chosen worker. High ability workers in pursuit of higher wages would want to signal employers about their greater productivity. Also, employers would benefit from such a signal by selecting more productive workers and avoiding the cost of hiring and training workers who may potentially be fired. Education might be a credible signal if high ability workers can complete education with less effort (lower cost) than low ability workers. If education is used as a signal of differentiating high ability workers from low ability workers, employers and employees benefit from this information; however, in this model there is no other benefit from public spending on education. Even when there is a positive relationship between workers' productivity and spending on education, we find some overinvestment in education (Milgrom and Roberts, 1992).

Signaling has been used to explain a variety of other strategic decisions (Riley, 2001). The qual ity of the product that is known to the producer but not to the potential customer can be signaled by using price, ADVERTISING, guarantees or warranties, and so on. For example, a high qual ity producer may use a low price or uninforma tive advertising to signal high quality if his cost of using this signal is lower than that of a low quality producer (Nelson, 1974). Milgrom and Roberts (1982) show that when a firm that is entering the market does not know the costs of the established firm in the market, the estab lished firm may be able to deter ENTRY by charging a lower price because the reduced price signals to potential entrants that the costs of the established firm are so low that as entry occurs the entrant will experience a loss (*see* LIMIT PRICING). Milgrom and Roberts (1986) show that for goods that are purchased often, both price and advertising could be employed to provide signals to consumers. For these goods, high quality firms have an incentive to set a low

price expecting to earn profits on future pur chases. On the other hand, durable goods should enter the market at a high price and as Bagwell and Riordan (1991) show, the price should de crease with time as information about the quality of the good spreads. In financial markets man agers have significantly more information about their company's performance and prospects than their investors. Signaling has been employed to analyze financial decisions which have the ability to influence what investors believe about the firm's prospects. For example, Ross (1977) showed that a choice of a company to increase its debt–equity ratio provided a signal for invest ors that its equity shares were more valuable.

See also *adverse selection; firm financial struc ture*

Bibliography

Bagwell, K. and Riordan, M. K. (1991). High and declin ing prices signal product quality. *American Economic Review*, **81**, 224 39.

Milgrom, P. and Roberts, J. (1982). Limit pricing and entry under incomplete information: An equilibrium analysis. *Econometrica*, **50**, 443 59.

Milgrom, P. and Roberts, J. (1986). Price and advertising signals of product quality. *Journal of Political Economy*, **94**, 796 821.

Milgrom, P. and Roberts, J. (1992). *Economics, Organiza tion and Management*. Englewood Cliffs, NJ: Prentice-Hall.

Nelson, P. (1974). Advertising as information. *Journal of Political Economy*, **82**, 729 54.

Riley, J. G. (2001). Silver signals: Twenty-five years of screening and signaling. *Journal of Economic Literature*, 432 78.

Ross, S. (1977). The determination of financial structure: The incentive signaling approach. *Bell Journal of Eco nomics*, **8**, 23 40.

Spence, A. M. (1974). *Market Signaling: Informational Transfer in Hiring and Related Screening Processes*. Cam bridge, MA: Harvard University Press.

simultaneous equations bias

Robert E. McAuliffe

This refers to the bias from using LINEAR RE GRESSION to estimate the parameters of an equation that is part of a system of simultaneous

equations. For example, when market prices and quantities sold are determined by both supply and demand, direct estimates of the demand or supply curves by linear regression will be biased because one of the basic assumptions regarding the ERROR TERMS IN REGRESSION is violated.

To see this, consider the supply and demand for any good below:

$$Q_t^d = \alpha_1 + \alpha_2 P + \varepsilon_t^d \qquad (1)$$

$$Q_t^s = \beta_1 + \beta_2 P + \varepsilon_t^s \qquad (2)$$

$$Q_t^d = Q_t^s \qquad (3)$$

where Q is the quantity supplied or demanded, P is the price of the product, and the ε_t^i are random disturbances to demand and supply. Equations (1)–(3) are referred to as *structural equations* which reflect specific features of the market. These equations are mutually dependent so random variations in demand or supply will cause random variations in price. But if demand is randomly high (or low) in a given period, the price will be higher (lower) in that period as well, and this means there is a correlation between one of the right hand variables in the regression (price) and the error term which biases linear regression estimates. The same problem occurs with estimates of the supply curve or with estimates of any simultaneous relationship where variables on the right hand side of the regression are not exogenous.

To see why the estimates from linear regression are biased, substitute (1) and (2) into (3) above and solve for price:

$$P = \frac{(\beta_1 + \varepsilon_t^s - \alpha_1 - \varepsilon_t^d)}{\beta_2 - \alpha_2} \qquad (4)$$

The linear regression estimator for the slope of the demand curve, α_2, α_2^L, is calculated as:

$$\alpha_2^L = \frac{\text{cov}(Q^d, P)}{\text{var}(P)} \qquad (5)$$

where $\text{cov}(Q^d, P)$ is the COVARIANCE between the quantity demanded and price, and $\text{var}(P)$ is the VARIANCE of the price data. For the linear regression estimator to be unbiased, there must be no correlation (or covariance) between the

right hand variables in the regression (P in this example) and the error term, ε_t^d. To check this assumption, take the covariance of the demand curve (equation (1) above) with respect to P. This yields:

$$\text{cov}(Q^d, P) = \alpha_2 \text{var}(P) + \text{cov}(P, \varepsilon_t^d) \qquad (6)$$

Dividing both sides by the variance of P and solving for α_2 provides the linear regression estimator of α_2:

$$\alpha_2^L = \frac{\text{cov}(Q^d, P)}{\text{var}(P)} - \frac{\text{cov}(P, \varepsilon_t^d)}{\text{var}(P)} \qquad (7)$$

The first term on the right in equation (7) is the correct estimator of α_2 but the second term, which represents the covariance between the error term and the price variable, must be zero for the linear regression estimator to be unbiased. Since the solution for price in equation (4) shows that variations in the error term will cause variations in price, the covariance between P and ε_t^d will not be zero as needed, and thus the estimator is biased. The degree to which the linear regression estimator is biased will depend on the size of the covariance between price and the error term, the last term in equation (7).

There are several methods for obtaining unbiased estimators of the parameters in simultaneous equations systems (Greene, 2002; Maddala, 2001; *see* ESTIMATING DEMAND), but it is important for researchers to know that linear regression methods cannot be used in general to estimate demand or supply equations directly.

Bibliography

Greene, W. H. (2002). *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.

Gujarati, D. N. (2002). *Basic Econometrics*, 4th edn. New York: McGraw-Hill Irwin.

Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.

standard deviation

Robert E. McAuliffe

The standard deviation of a random variable is the square root of its VARIANCE and it measures dispersion around the EXPECTED VALUE, or

mean of that random variable. For a discrete random variable, X, which has specific outcomes denoted by X_i and where the probability distribution of the values of X is given by $f(x)$, then the standard deviation of X, σ, is calculated as

$$\sigma = \sqrt{\sum f(x) \times (X_i - E(X))^2} \qquad (1)$$

where $E(X)$ is the expected value of the random variable, X, and Σ denotes the summation over all possible values of X. Each squared deviation of the outcome, X_i, from the expected value, $E(X)$, is weighted by its probability, $f(x)$.

A manager who wishes to calculate the standard deviation for a sample of T observations would estimate the standard deviation using

$$s = \sqrt{\sum \frac{1}{T-1} \times (X_i - \bar{X})^2} \qquad (2)$$

where $\bar{X}$ is the sample mean. Note that the squared deviations are divided by $T - 1$ and not by T because a *degree of freedom* has been lost. Some of the sample information has already been used to calculate the sample mean, and given this calculation only $T - 1$ of the observations are independent (see Maddala, 2001; Greene, 2002). Most numerical software programs and spreadsheets will calculate the sample standard deviation for a set of observations.

Since more risk is associated with greater dispersion in the possible outcomes, the standard deviation can be used as a measure of risk. The COEFFICIENT OF VARIATION, one measure of risk per dollar of expected return, uses the standard deviation as a measure of risk. The standard error of an estimated coefficient can be used for hypothesis testing to determine if an explanatory variable in a LINEAR REGRESSION has a significant effect on the dependent variable. Such a test would be useful to managers who may wish to determine if a recent change in policy (ADVERTISING, pricing, etc.) had significant effects on sales, profits, or some other variable of interest.

See also *t statistic*

Bibliography

Greene, W. H. (2002). *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.
Gujarati, D. N. (2002). *Basic Econometrics*, 4th edn. New York: McGraw-Hill Irwin.
Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.

standard error of estimate

Robert E. McAuliffe

The standard error of estimate (SEE) is the square root of the VARIANCE of the residuals from a LINEAR REGRESSION. This information is crucial because no standard errors or CONFIDENCE INTERVALS can be calculated without it. For a simple linear regression with T observations and one explanatory variable, X, the population relation between the dependent variable, Y, and X would be

$$Y_i = \beta_0 + \beta_1 \times X_i + v_i \qquad (1)$$

where the Y_i are observations on the *dependent* variable, the X_i are observations on the *independent* or *explanatory* variable, and the v_i are the population error terms. The population error terms are assumed to have zero mean and variance σ^2. Since the true population values of β_0, β_1, and v_i are not known, they must be estimated. If $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimators of the population values, then the estimated residuals are given by:

$$\hat{v}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 \times X_i \qquad (2)$$

The estimator for the population variance, σ^2, is s^2 and is calculated as:

$$s^2 = \frac{1}{T-2} \sum_{i}^{T} \hat{v}_i^2 \qquad (3)$$

and the SEE for the regression is simply the square root of equation (3) above. Note that the sum of the squared residuals is divided by $T - 2$ rather than by T because two degrees of freedom are lost estimating β_0 and β_1 (see Maddala, 2001; Greene, 2002). In a multiple regression with T observations, k explanatory variables, and a constant term, the estimated variance in equation (3) would be divided by $T - k - 1$.

The SEE is also useful as a guide in forecasting. Larger values of the SEE imply larger fore

casting errors, all else equal, and the SEE can be compared to the size of the dependent variable to gauge the likely size of forecast errors.

Bibliography

Greene, W. H. (2002). *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.

Gujarati, D. N. (2002). *Basic Econometrics*, 4th edn. New York: McGraw-Hill Irwin.

Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.

standard error of the coefficient

Robert E. McAuliffe

In a LINEAR REGRESSION this is the estimated standard error which is used to test hypotheses regarding the estimated coefficients. In a simple linear regression with T observations and one explanatory variable, the population regression relation is:

$$Y_i = \beta_0 + \beta_1 \times X_i + v_i \qquad (1)$$

where the Y_i are observations on the *dependent* variable, the X_i are observations on the *independent* or *explanatory* variable, and the v_i are the population error terms. It can be shown that the VARIANCE of the ordinary least squares estimator of β_1, $\hat{\beta}_1$, is

$$\text{var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2} \qquad (2)$$

where σ^2 is the population variance (see Maddala, 2001; Greene, 2002). Since the population variance is generally unknown, it is replaced by the estimator, s^2, which is calculated as:

$$s^2 = \frac{1}{T-2} \sum_{i=1}^{T} \hat{v}_i^2 \qquad (3)$$

(*see* STANDARD ERROR OF ESTIMATE). The estimated standard error for the coefficient $\hat{\beta}_1$ is then:

$$s\hat{\beta}_1 = \sqrt{\frac{s^2}{\sum (X_i - X)^2}} \qquad (4)$$

This value is reported by most statistical software programs as part of the standard regression output, often with the appropriate t statistics to test the null hypothesis that the coefficient is not significantly different from zero (*see* T STATISTIC). Note that the standard error for the coefficient above depends on the estimated variance of the error terms, the variance of the explanatory variable, and the number of observations corrected for degrees of freedom. Intuitively this means that an estimated coefficient is more reliable the greater the range (variance) of X values in the sample because the data have more information. In a multiple regression with k explanatory variables, a constant term, and T observations, the Xs in equation (4) would be matrices and the sum of squared residuals in equation (3) would be divided by $T - k - 1$ (see Maddala, 2001; Greene, 2002).

Bibliography

Greene, W. H. (2002). *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.

Gujarati, D. N. (2002). *Basic Econometrics*, 4th edn. New York: McGraw-Hill Irwin.

Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.

stochastic dominance

Roger Tutterow

Situations involving risk are often modeled as a probability distribution across a set of possible outcomes. In this environment, optimal decisions depend not only upon the outcomes and their associated probabilities, but also upon the decision maker's attitude toward risk.

Early work on choice under UNCERTAINTY centered upon characterizing agents' preferences over statistical moments of a probability distribution. In particular, portfolio theory and asset valuation models frequently assume preferences over the first two moments, mean (*see* EXPECTED VALUE; VARIANCE). Generally, one distribution is said to be preferred to another distribution if the former has either the same mean and lower variance or a higher mean and

the same variance. Markowitz (1950) is a stand ard reference on the application of this frame work to portfolio theory.

While the mean–variance framework is intui tively appealing, it is generally inconsistent with the premier theoretical paradigm for choice under uncertainty – von Neumann and Morgen stern's expected utility hypothesis. In expected utility theory, a rational agent's preferences over uncertain outcomes may be represented by a continuous utility function such that the agent will evaluate lotteries in terms of expected util ity. Characterizing an agent's preference in terms of mean and variance is only consistent with expected utility for a few restrictive cases, those in which the utility function is quadratic or the outcomes are normally distributed. In con trast, the stochastic dominance ordering is always consistent with expected UTILITY MAXIMIZATION. Further, it incorporates the entire probability distribution rather than a few statistical moments.

Let $F(z)$ and $G(z)$ denote two cumulative distribution functions over the set of outcomes z. Intuitively, $F(z^*)$ denotes the probability of obtaining an outcome at least as large as z^*. Then we say $F(z)$ first degree stochastic dominates $G(z)$ if $F(z) \leq G(z)$ for all z with the inequality holding strictly for some z. Quirk and Saposnik (1962) showed that the stochastic dominating distribution will be preferred by all expected utility maximizing agents who prefer more income to less. Thus, first degree stochastic dominance makes no assumption about the deci sion maker's attitude toward risk.

By making further assumptions about the agent's preference, the stochastic dominance ordering may be generalized to higher degrees. Hadar and Russell (1969) and Hanoch and Levy (1969) introduced the notion of second degree stochastic dominance. Formally, $F(z)$ second degree stochastic dominates $G(z)$ if

$$\int_{\infty}^{z} F(x)dx \leq \int_{\infty}^{z} G(x)dx$$

for all z with the inequality holding strictly for some z. Second degree stochastic dominance is equivalent to expected utility maximization for all agents with increasing, concave utility func

tions, i.e., those agents who display RISK AVER SION.

Stochastic dominance may be extended to an nth degree ordering. For a review of these ex tensions as well as the application of stochastic dominance to financial valuation models, the reader is referred to Ingersoll (1987) and the citations contained therein.

Bibliography

Hadar, J. and Russell, W. R. (1969). Rules for ordering uncertain prospects. *American Economic Review*, **59**, 25 34.

Hanoch, G. and Levy, C. (1969). Efficiency analysis of choices involving risk. *Review of Economic Studies*, **36**, 335 46.

Ingersoll, J. E., Jr. (1987). *Theory of Financial Decision Making*. Ottowa: Rowman and Littlefield.

Markowitz, H. M. (1950). *Portfolio Selection*. Cowles Foundation Monograph, 16. New Haven, CT: Yale University Press.

Quirk, J. P. and Saposnik, R. (1962). Admissibility and measurable utility functions. *Review of Economic Stud ies*, **29**, 140 6.

von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Prince ton University Press.

stockholders

John Edmunds and Roberto Bonifaz

Financial resources for the firm come from lenders and stockholders. Lenders put money into the firm for stated periods of time, and require to be paid stipulated amounts of interest. When the terms of the loan agreement call for repayment to occur, lenders can legally demand to be repaid. Stockholders put money into the firm without any expectation or legal right to be repaid at any particular time in the future. They may expect to earn dividends on the money they invest, but the money is permanently invested in the business, so stockholders can only realize a lump sum recovery of their investment by sell ing their shares to other investors. The firm does not have a legal obligation to redeem the shares which the founding stockholders bought at the time the firm was created. The stockholders, therefore, are residual claimants on the income

stream of the firm. Senior claimants are employ ees, trade creditors, tax authorities, and lenders, including banks and bondholders. The stock holders are in the most junior position, and consequently face the greatest risk of being wiped out if the firm fails.

Stockholders have control, in theory at least, over the firm. As long as the firm is in compli ance with the terms of its agreements with senior claimants, the stockholders have the right to name the board of directors, which in turn names the line managers who run the day to day affairs of the firm. They can benefit if the firm is very profitable, or if another firm tries to buy enough shares to acquire control. At the same time, however, the interests of the stock holders may not be the same as those of the firm's managers (*see* PRINCIPAL–AGENT PROBLEM).

It is important to note that the firm's man agers may develop goals distinctly different from those of the stockholders (because of the princi pal–agent problem). Herbert Simon (1959) has emphasized that managers tend to behave in ways that are "satisfactory" to the stockholders rather than profit maximizing (*see* PROFIT MAXIMIZATION). Instead of managers maxi mizing long run profits, they maximize individ ual goals subject to the constraint of satisfying stockholders.

Stockholders can benefit if the firm's man agers optimize its CAPITAL structure, by repur chasing shares in the market and subsequently holding them in the treasury, or by borrowing money and paying a large one time cash divi dend. Stockholders are also better served by managers when CONTRACTS and managerial rewards align the managers' interests with those of the stockholders. For example, when managers' rewards are in the form of stock options, they have greater incentives to maxi mize profits and raise the value of their shares, which is in the stockholders' interest as well.

Bibliography

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Milgrom, P. and Roberts, J. (1992). *Economics, Organiza tion and Management*. Englewood Cliffs, NJ: Prentice-Hall.

Simon, H. A. (1959). Theories of decision making in economics. *American Economic Review*, **49**, 253 83.

strategic behavior

Robert E. McAuliffe

Firms behave strategically when they engage in activities intended to change their rivals' expect ations. A firm may reduce prices or significantly increase ADVERTISING expenditures when threatened by ENTRY in an attempt to prevent entry. If these are CREDIBLE STRATEGIES, then entry may be deterred. In addition, the firm may establish a reputation for aggressive behavior to send a signal to other potential en trants.

A variety of strategic options are available to firms, such as investing in EXCESS CAPACITY, engaging in research and development, pricing decisions, PRODUCT DIFFERENTIATION, QUALITY choice, and PRODUCT PROLIFER ATION. When the US Federal Trade Commis sion charged the ready to eat breakfast cereals firms with antitrust violations, one of the allega tions made was that the existing firms intro duced new brands to prevent entry. By saturating the market with new brands, there was no room left for new entry despite the high profits earned by the existing firms (see Schma lensee, 1978).

Bibliography

Besanko, D., Dranove, D., Shanley, M., and Shaefer, S. (2003). *Economics of Strategy*, 2nd edn. New York: John Wiley.

Carlton, D. W. and Perloff, J. M. (2000). *Modern Indus trial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Nagle, T. T. and Holden, R. K. (2002). *The Strategy and Tactics of Pricing: A Guide to Profitable Decision Making*, 3rd edn. Upper Saddle River, NJ: Prentice-Hall.

Schmalensee, R. (1978). Entry deterrence in the ready-to-eat breakfast cereal industry. *Bell Journal of Economics*, **9**, 305 27.

structure-conduct-performance paradigm

Gilbert Becker

This basic model from the field of industrial organization economics is used both as a vehicle for industry analysis and to provide a framework for public policy discussion. By examining the three elements of the paradigm as they pertain to a specific industry, a firm's managers are better able to evaluate the competitive nature and process of that industry. Knowledge of the elements of MARKET STRUCTURE such as the condition of ENTRY into the market and the number of firms already within, as well as the prevailing forms of strategic rivalry, is shown by Porter (1980) and others as being essential to building a successful long term business strategy. Researchers have explored numerous industries in the manufacturing, retail, service, and other sectors using this paradigm (see Adams and Brock, 2001; Deutsch, 1998).

Historically, the paradigm has primarily been used by economists and public policy officials for the study of industrial policy questions. Considerable controversy has evolved as a result of the two distinct views which the paradigm supports. The early development of the paradigm by Bain (1951, 1956) and others presented the traditional view that market structure was causally responsible for market conduct and ultimately for MARKET PERFORMANCE. This position, known as the **structuralist view**, maintains that noncompetitive (e.g., MONOPOLY and OLIGOPOLY) market structures lead to undesirable forms of rivalry and poor market performance relative to their more competitive counterparts (*see* MONOPOLISTIC COMPETITION; PERFECT COMPETITION). Bain's argument that markets having high CONCENTRATION INDICES were environments that were conducive to price fixing and other forms of COLLUSION has come to be called the **concentration collusion hypothesis**. In addition, the structuralist view holds that high BARRIERS TO ENTRY foster excess ECONOMIC PROFIT, and that these structural conditions result in technical inefficiency by producers and a misallocation of society's resources (*see* EFFICIENCY).

The economic analysis resulting from this view yields several conclusions that may assist managers in evaluating a market in which they wish to compete. First, the level of interdependence in rivals' conduct is related to the number of rivals and their size. Pricing strategies are more interdependent in markets with fewer rivals and high concentration levels. For example, in some markets PRICE LEADERSHIP may be in effect. Also, price SIGNALING, overt collusion, and even tacit awareness of acceptable forms of rivalry are more likely since fewness in numbers facilitates communication and diminishes cheating on commonly accepted industry goals. A lack of PRODUCT DIFFERENTIATION and a similarity of cost conditions across rivals are two other structural conditions that are conducive to greater uniformity in pricing strategies among the rivals. Finally, strategies including ADVERTISING expenditures, brand proliferation (*see* PRODUCT PROLIFERATION), control of shelf space, and the creation of excess industry capacity have been cited as forms of conduct that may evolve in highly concentrated industries in an effort to forestall entry (see LIMIT PRICING).

Considerable empirical investigation of these arguments has been performed. Bain (1956) demonstrated that both high industry concentration and high entry barriers were associated with significantly higher profit rates. Following Bain, a number of empirical investigations using LINEAR REGRESSION to examine a cross section of several industries were performed over the next two decades. A variety of market concentration indices were examined for their effect on performance variables such as price–cost margins and rates of return on owners' equity. The theoretical underpinning of many of these studies was fortified by Cowling and Waterson (1976). Their work showed that a strong theoretical link exists between the HERFINDAHL HIRSCHMAN INDEX (HHI), which measures industry concentration, and the LERNER INDEX (L), which measures performance. Specifically, it was demonstrated that

$$L = \frac{\text{HHI}}{|\varepsilon_p|}$$

where ε_p measures the price ELASTICITY of demand. Weiss (1974), in a summary of over 40 of the early studies, concludes that this evidence shows support for the concentration–collusion

hypothesis. Several of these studies also in cluded measures of barriers to entry arising from advertising expenditures, CAPITAL re quirements, and ECONOMIES OF SCALE. These studies found additional support for the structuralist view. Taken as a whole, the studies have been used to support antitrust policy (*see* ANTITRUST POLICY (US)) which calls for careful scrutiny of monopoly and oligopoly be havior and governmental action to break up those industries having highly concentrated market structures. In addition, the use of HORI ZONTAL MERGER GUIDELINES to prevent the occurrence of increasing concentration has also been justified by this analysis.

Controversy about the traditional view arises from a competing version of the paradigm which argues that the causal linkage between structure, conduct, and performance is much more com plex, especially in that market structure is not exogenously determined. Profitability, TECH NICAL EFFICIENCY, and other market per formance variables, coupled with business strategy, can and do alter market structure. At the center of this view is the **efficiency hypoth esis**, which holds that market structure is deter mined by a competitive process that forces firms to achieve MINIMUM EFFICIENT SCALE (MES) in order to survive. As such, minimum firm size relative to overall market demand dic tates that some industries be more concentrated than others if society is to achieve maximum benefits from cost efficient production.

In one early but important empirical study in support of this alternative view, Demsetz (1973) examines the impact of industry concentration on the profit rates of firms of different sizes. If the collusion hypothesis were correct, small firms would also be expected to benefit by following their larger rivals to the new higher prices. His findings demonstrated that as indus try concentration rose, the profit rates for large firms also tended to rise, but that no such tendency was exhibited for smaller firms. The concentration–collusion hypothesis was thus rejected in favor of the explanation that the greater profits of the larger firms in concentrated industries were due to their efficiency.

Numerous measurement problems (see the entry on market performance) have also led to criticism of the early studies of the structuralist view. One problem in particular is that the use of accounting data to measure profits or the Lerner index yields flawed estimates of market perform ance (*see* ACCOUNTING PROFIT). Moreover, these earlier studies did not examine the impact of concentration and barriers on performance in the long run. Some studies now show that the high profits of concentrated industries tend to decline over time. In a broad survey of the more recent literature, Schmalensee (1989) finds that the support for the concentration–profits link is weak. In addition he finds some support for the efficiency hypothesis in recent studies which show that concentration is positively related to various measures of efficient scale.

Some of the recent literature, in the spirit of the structuralist view, has turned to econometric modeling of specific industries in an effort to establish the extent of market power existing in an industry. These models investigate the effects of structure and conduct upon prices and other market outcomes. Bresnahan's survey (1989) of this literature concludes that the evidence indi cates that some concentrated markets hold con siderable market power. Since other studies have also shown that industry concentration levels cannot be fully explained by the existing tech nology, the controversy between the two views of the paradigm continues.

See also *EU competition policy, 2004; EU merger guidelines, 2004; EU merger policy, 2004; market power; merger guidelines, 1992–7*

Bibliography

Adams, W. and Brock, J. (2001). *The Structure of Ameri can Industry*, 10th edn. Upper Saddle River, NJ: Pren tice-Hall.

Bain, J. S. (1951). Relation of profit rate to industry concentration: American manufacturing, 1936 1940. *Quarterly Journal of Economics*, **65**, 293 324.

Bain, J. S. (1956). *Barriers to New Competition*. Cam bridge, MA: Harvard University Press.

Bresnahan, T. (1989). Industries with market power. In R. Schmalensee and R. D. Willig (eds.), *Handbook of Industrial Organization*. New York: North-Holland.

Cowling, K. and Waterson, M. (1976). Price cost margins and market structure. *Economica*, **43**, 267 74.

Demsetz, H. (1973). Industry structure, market rivalry and public policy. *Journal of Law and Economics*, **16**, 1 9.

Deutsch, L. (1998). *Industry Studies*, 2nd edn. Armonk, NY: M. E. Sharpe.

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.

Schmalensee, R. (1989). Interindustry studies of structure and performance. In R. Schmalensee and R. D. Willig (eds.), *Handbook of Industrial Organization*. New York: North-Holland.

Shepherd, W. and Shepherd, J. (2004). *The Economics of Industrial Organization*, 5th edn. Long Grove: Waveland Press.

Waldman, D. and Jensen, E. (2001). *Industrial Organization: Theory and Practice*, 2nd edn. Boston: Addison-Wesley.

Weiss, L. (1974). The concentration profits relationship and antitrust. In H. Goldschmid, H. Mann, and J. Weston (eds.), *Industrial Concentration: The New Learning*. Boston: Little, Brown.

substitutes

Robert E. McAuliffe

When products or inputs are substitutes, one may be used instead of the other(s) to satisfy demand or meet production requirements. Products which are substitutes in consumption should have a large, positive cross elasticity of demand (*see* CROSS ELASTICITIES), where if the price of butter rises, for example, it causes an increase in the quantity demanded of margarine. Inputs in production may also be substitutes depending on the technology of production employed. A rise in the cost of skilled labor may increase the demand for unskilled labor or for labor saving machinery if technology permits.

Bibliography

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

substitution effect

Robert E. McAuliffe

A change in the price of a product affects consumer demand through the INCOME EFFECT and substitution effect. If the price of a product rises, the real price of the product increases (*see* REAL PRICES) and the consumer's BUDGET CONSTRAINT is more binding. The increase in the relative price of butter, for example, will cause the consumer to substitute other products, such as margarine, for butter, and the consumer will move along her indifference curve (if possible) substituting margarine for butter (*see* INDIFFERENCE CURVES). The substitution effect refers to this movement along the consumer's indifference curve. The income effect from the price change may force the consumer to a new indifference curve if consumers spend a significant portion of their income on this good. This occurs because a price increase reduces the consumer's REAL INCOME and leaves the consumer worse off (see Douglas, 1992; Shughart, Chappell, and Cottle, 1994).

Bibliography

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Shughart, W. F., Chappell, W. F., and Cottle, R. L. (1994). *Modern Managerial Economics*. Cincinnati, OH: South-Western Publishing.

sunk costs

Robert E. McAuliffe

Sunk costs are those costs which cannot be recovered by the firm. As with FIXED COSTS, these costs do not vary with output produced and, as Tirole (1988) suggests, the differences between the two are matters of degree. Consider a firm that rents a machine for $100,000 per month which has an economic life of 2 years and costs $2 million to purchase. The $100,000 monthly rental fee is a fixed cost each month to the firm. Suppose instead that the firm decided to purchase the machine and, once purchased, the resale value of the machine was zero. Now the $2 million purchase price is a sunk cost to the firm because the asset has no alternative uses (its opportunity cost is zero; *see* OPPORTUNITY COSTS). If the firm could sell the machine for

$1 million, then the sunk cost would be only $1 million, that portion of the original price which is not recoverable. The portion of any asset's purchase cost which is sunk will depend upon the resale market for that asset. Assets which are very specialized and unique to an industry or for which resale markets do not exist or are "thin" will have a greater percentage of their purchase cost "sunk" (*see* ASSET SPECIFICITY).

Sunk costs play an important role in determining MARKET STRUCTURE because they represent a barrier to entry to new firms (*see* BARRIERS TO ENTRY; ENTRY). Once incurred, sunk costs should have no effect on the pricing and output decisions of established firms and, if necessary, established firms would produce output as long as AVERAGE VARIABLE COSTS were covered. This is not the case for a potential entrant who must pay these costs to enter the industry and must expect to cover them if entry is to succeed. An asymmetry is thus created between the established firms, for whom these costs are irrelevant, and potential entrants, who must include them in the entry decision. Sunk costs act as a barrier to EXIT once firms are in an industry and may affect the intensity of COMPETITION between established firms, particularly if the industry is declining (*see* DECLINING INDUSTRY; INDUSTRY LIFE CYCLE).

Baumol, Panzar, and Willig (1982) argued that when entry and exit were absolutely free and costless, the market was perfectly contestable (*see* CONTESTABLE MARKETS). Free, costless entry and exit require that entrants have no sunk costs (even if there are fixed costs), enabling them to quickly enter the industry, undercut the established firms, and exit before the incumbents could retaliate. In these markets the threat of entry would discipline established firms (even a natural monopolist; *see* NATURAL MONOPOLY) to the point where they earned no ECONOMIC PROFIT. But Weitzman (1983) argued that there are no purely fixed costs and, theoretically, ECONOMIES OF SCALE cannot exist without sunk costs. Both fixed costs and sunk costs imply some commitment by the firm for some period of time. For fixed costs the commitment lasts for the duration of the SHORT RUN, while sunk costs commit the firm for a longer period of time (see Tirole, 1988). In replying to their critics, Baumol et al. usefully distinguished economic sunk costs from technological sunk costs. Although a firm might need a period of time to physically produce the goods or services for the market, an entrant could enter into conditional CONTRACTS with buyers and sell its output before actually producing the product. In such a case, the entrant would not incur the sunk costs of production unless enough contracts were sold and, once sold, the existing firm(s) would be subject to hit and run entry.

Sutton (1991) extended the notion of sunk costs by suggesting that there may be endogenous sunk costs in an industry in addition to exogenous sunk costs. He developed a two stage model where in the first stage, if a firm chose to enter the industry, it would incur fixed setup costs to produce output, such as acquiring a MINIMUM EFFICIENT SCALE (MES) plant and incurring ADVERTISING and research and development expenses to establish the product(s). In the second stage, those firms in the industry would compete for profits through price or quantity competition. Sutton considered the fixed costs of acquiring an MES plant as exogenous sunk costs, while the advertising and research and development expenditures were endogenous sunk costs because firms could adjust these costs in stage 1 to improve their competitive position in stage 2. In industries where exogenous sunk costs were most significant, the industry structure would become less concentrated as the market grew, but in those industries where endogenous sunk costs were more important, a concentrated industry structure arose even as market size increased. He found that industries with high concentration and high advertising and research and development expenditures in one country were also highly concentrated in other countries. He attributed this result to a competitive escalation of these endogenous sunk expenditures which kept industry concentration levels high and prevented entry from having a significant effect on industry structure.

Bibliography

Baumol, W. J., Panzar, J. C., and Willig, R. D. (1982). *Contestable Markets and the Theory of Industry Structure*. San Diego: Harcourt Brace Jovanovich.

Baumol, W. J., Panzar, J. C., and Willig, R. D. (1983). Contestable markets: An uprising in the theory of industry structure: Reply. *American Economic Review*, **73**, 491 6.

Schwartz, M. and Reynolds, R. J. (1983). Contestable markets: An uprising in the theory of industry structure: Comment. *American Economic Review*, **73**, 488 90.

Shepherd, W. G. (1984). "Contestability" vs. competition. *American Economic Review*, **74**, 572 87.

Sutton, J. (1991). *Sunk Costs and Market Structure*. Cambridge, MA: MIT Press.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

Weitzman, M. L. (1983). Contestable markets: An uprising in the theory of industry structure: Comment. *American Economic Review*, **73**, 486 7.

survivor principle

Govind Hariharan

In order to determine the ability of any firm in an industry to successfully compete in the market, cost functions are typically estimated (*see* COM PETITION; SHORT RUN COST CURVES). How ever, for quick and easy casual information on the structure of cost and ECONOMIES OF SCALE in an industry, a popular approach is to look at firms of different scales or sizes over time to see which sizes survived. If it is observed that small firms in one industry tended to go out of business or were taken over while medium and large firms continued to survive, it leads to the possibility that in this industry the small firms had higher per unit costs (*see* AVERAGE TOTAL COST), and hence could not compete with the medium and large firms. In this example, the smallest size or scale of operations at which firms become competitive in this industry and have the potential to survive is then the medium sized firm. In order to survive, smaller firms must either merge or increase their scale of op eration. Such an analysis based on survival trends in an industry was made popular by Nobel economist George Stigler (Stigler, 1958), and is referred to as survivor principle/ analysis. All that is needed for such an analysis is to break up the firms in an industry into differ ent size classes and observe which size class or classes seem to be growing and which are declin

ing. In the physician medical practice industry, for example, it was found that between 1965 and 1980, practices with one or two physicians had the largest market share, but their share was declining throughout the period (Mardel and Zuckerman, 1985). This suggests that in this industry small group practices may no longer be very efficient and one would expect a gradual movement toward larger sized practices. In an other study, Blair and Vogel (1978) looked at the health insurance industry between 1958 and 1973 and observed that the larger sized firms seemed to be surviving, and hence were likely to be more efficient. In many of these studies, including a study on manufacturing (Weiss, 1964), a wide range of sizes was found to be able to survive, suggesting that a large number of relatively smaller firms could survive provid ing some semblance of a competitive market (*see* MINIMUM EFFICIENT SCALE; PERFECT COMPETITION).

Bibliography

Blair, R. D. and Vogel, R. J. (1978). A survivor analysis of commercial health insurers. *Journal of Business*, **51**, 521 30.

Mardel, W. D. and Zuckerman, S. (1985). Competition and medical groups: A survivor analysis. *Journal of Health Economics*, **4**, 167.

Stigler, G. J. (1958). The economies of scale. *Journal of Law and Economics*, **1**, 54 71.

Weiss, L. (1964). The survivor techniques and the extent of suboptimal capacity. *Journal of Political Economy*, **72**, 246 61.

switching costs

Robert E. McAuliffe

Switching costs occur when a potential con sumer of a good finds a better product but does not purchase it because of higher costs that will arise. For example, if a household purchased a new car a few months ago and a new and better model was just announced, it would be unlikely that the household would purchase the new model. Switching costs can arise in a variety of situations – when purchasing software or in selecting a computer, for example. In any case, where consumers must invest time and energy to

buy or use a product, some switching costs will be incurred.

Typically, switching costs are low and do not interfere with economic EFFICIENCY. However, there are potential cases – particularly regarding purchases of high tech products – when switching costs could be important and prevent consumption of the "better" good. These circumstances were thought to arise when consumers purchased a newly introduced good and became locked in to that particular product (*see* LOCK IN). This meant that if a better product subsequently appeared, consumers would not purchase it because of the value the original product had with users. For a time this fueled the race on the Internet to be the first supplier of a given product. It also suggested the possibility that existing standards in the market were not the best – particularly when there were NETWORK EXTERNALITIES.

As subsequent research has shown (see Leibowitz and Margolis, 1990, 1995; Leibowitz, 2002), market success is not so easily earned and those products that succeed do so because they are better. Despite the intriguing possibility, switching costs have not interfered with economic efficiency to date.

Bibliography

Leibowitz, S. J. (2002). *Rethinking the Network Economy*. New York: American Management Association.

Leibowitz, S. J. and Margolis, S. E. (1990). The fable of the keys. *Journal of Law and Economics*, **33**, 1 26.

Leibowitz, S. J. and Margolis, S. E. (1995). Path dependence, lock-in, and history. *Journal of Law, Economics and Organization*, **11**, 205 26.

Shapiro, C. and Varian, H. R. (1999). *Information Rules*. Boston: Harvard Business School Press.

synergies

Gilbert Becker

Synergies is a business term used to identify potential cost savings or other gains resulting from the simultaneous production of more than one good or, more commonly, from the combination of two firms through a merger. As such, the concept is related to ECONOMIES OF SCALE and ECONOMIES OF SCOPE and may include

both in its measure, but it is not limited to these as it is generally not a clearly defined term.

Sirower (1997) developed a model of four cornerstones necessary to achieving synergies. These are:

1 a strategic vision for the newly created firm;
2 a mechanism for rewards and decision making, and for solving corporate culture differences within the two newly merged firms;
3 an operating strategy for the new firm to better compete with its rivals; and
4 the physical integration of sales, research and development, and other systems.

The improved ability to compete may result from acquiring access to a preferred distribution network, or by acquiring products that have complementary marketing or production needs. It may also occur as a result of expansion by the firm to a size needed to achieve MINIMUM EFFICIENT SCALE, or by enabling the removal of excess industry capacity while increasing the market share of the remaining firm (*see* EXCESS CAPACITY).

Sirower's study of major mergers and acquisitions during the 1980s cites corporate executives' predictions of expected synergies as the rationale offered for the payment of (sometimes significant) premiums above the prevailing market price for the acquired firm. Unfortunately, his work generates strong empirical evidence that in approximately two thirds of mergers valued at $100 million or more occurring during the 1980s, the merger resulted in a decline in profitability and shareholder value. Moreover, the size of the losses could be predicted, using LINEAR REGRESSION, and was proportional to the size of the premium offered by the acquiring firm.

He offers two possible explanations for these disappointing results. The first is the failure of management to carefully identify and develop each of the four cornerstones above. He concludes that these conditions often are not met, and more importantly, have not even been adequately examined prior to the merger. Failure to do so may increase costs in unexpected areas, or not allow expected cost savings to be achieved. The second reason involves an examination of

rivals' responses to actions taken by the firm. Here two additional preconditions necessary for achieving synergies from a proposed merger are:

1 the acquiring firm must be able to increase protection of its own markets as a result of the merger; and
2 the acquiring firm must be more able to expand into new markets or encroach on rivals' markets, without their responses being as great as they may have been without the merger.

The lack of either of these conditions may indicate that the merger did not generate MARKET POWER gains that were sufficient to deter rivals' potential destruction of any synergies achieved by the merger.

Evidence by Eckbo (1983), Lichtenberg and Siegel (1987), and others promotes the claim that mergers can successfully increase shareholder value and other measures of MARKET PERFORMANCE. The work of Sirower and others (e.g., Ravenscraft and Scherer, 1987) supports the caution that careful examination of synergies is necessary, prior to the merger, to increase the probability of its success.

Bibliography

Eckbo, E. (1983). Horizontal mergers, collusion and stockholder wealth. *Journal of Financial Economics*, **11**, 241 73.

Lichtenberg, F. and Siegel, D. (1987). Productivity and changes in ownership of manufacturing plants. *Brookings Papers on Economic Activity*, **3**, 643 83.

Ravenscraft, D. and Scherer, F. M. (1987). Mergers, selloffs and economic efficiency. Washington, DC: Brookings Papers on Economic Activity.

Sirower, M. (1997). *The Synergy Trap*. New York: Free Press.

T

t-statistic

Robert E. McAuliffe

Most statistical software programs report the t statistic along with other summary information as part of the routine output from a LINEAR REGRESSION. The t statistic is used for hy pothesis testing when the true population vari ance is unknown and is replaced by the estimated sample variance. For example, suppose a firm has TIME SERIES DATA on its sales and a man ager wants to determine if there is a significant relationship between its sales and GROSS DO MESTIC PRODUCT (GDP). The regression might be specified as:

$$\text{sales}_t = \beta_0 + \beta_1 \times \text{GDP}_t + v_t \qquad (1)$$

where v_t is a random error term. If there is no statistically significant relationship between sales and GDP, the estimated value for β_1 should be zero. But since β_1 is itself a random variable, even if its true value were zero we would observe a range of values for β_1 around zero. Given that any estimated value for β_1 is not likely to equal zero exactly, the t statistic is used to determine whether the estimated value is sufficiently dif ferent from zero so that random chance is un likely to be the cause and more likely that the true relationship is nonzero. For a regression with T observations, the t statistic is calculated as:

$$\frac{\beta_1 - \mu}{s_\beta} \qquad (2)$$

where μ is the hypothesized value for β_1 (the null hypothesis) and s_β is the estimated STANDARD ERROR OF THE COEFFICIENT for β_1, from the regression. The statistic above has a t distribu tion with $T - 2$ degrees of freedom because

sample information has been used to determine two coefficients, β_0 and β_1 (see Maddala, 2001; Greene, 2002). In the example here, the null hypothesis is that $\beta_1 = 0$, so the t statistic would then be

$$\frac{\beta_1}{s_\beta} \qquad (3)$$

which is simply the estimated value for the coef ficient divided by its standard error. If the sample size is large (more than 30 degrees of freedom), a 95 percent confidence interval is approximately β_1 plus or minus two standard deviations (*see* CONFIDENCE INTERVALS), so if the value in equation (3) above is greater than 2 in absolute value, then we would reject the null hypothesis that β_1 is equal to zero. In samples with fewer than 30 observations (or 30 degrees of freedom), critical values of the t distribution can be found in statistical tables (see Kmenta, 1997; Maddala, 2001). For example, if the estimated value of β_1 were 8 and its standard error were 2, then the value of the t statistic in equation (3) would be 4. When software programs provide t statistics as part of their standard regression output, the calculated values are based on equa tion (3) above where the null hypothesis is that the coefficient equals zero.

In the context of a multiple regression, a degree of freedom is lost for every coefficient which is estimated, including the constant term. Therefore if there are T observations, k explanatory variables, and a constant term, the test statistic will have $T - k - 1$ degrees of free dom.

Bibliography

Greene, W. H. (2002). *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.

Gujarati, D. N. (2002). *Basic Econometrics*, 4th edn. New York: McGraw-Hill Irwin.

Kmenta, J. (1997). *Elements of Econometrics*, 2nd edn. Ann Arbor: University of Michigan Press.

Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.

technical efficiency

Eduardo Ley

In physics, efficiency is the ratio of useful work done to total energy expended, and the same general idea is associated with the term in pro duction. Economists simply replace "useful work" by "outputs" and "energy" by "inputs." Technical efficiency means the adequate use of the available resources in order to obtain the maximum product. A productive activity is in efficient whenever it is possible to reduce the amount of some of its inputs without reducing the level of output, or whenever the output can be increased without using more inputs. Note that there is no reference to prices in this defin ition of efficiency, as opposed to allocative or economic EFFICIENCY.

Bibliography

Färe, R., Grosskopf, S., and Lovell, C. A. K. (1985). *The Measurement of Efficiency of Production*. Boston: Kluwer-Nijhoff.

time-series data

Robert E. McAuliffe

Time series data are defined as information gathered for specific variables which are ob served over a period of time. For example, a firm might have quarterly pricing and sales data available for 10 years which could be used to estimate the demand for its product. But a manager could also estimate demand if the firm had pricing and sales data available from 40 branches located across the country. The data in this second example are cross section data and are obtained by gathering information from dif ferent locations at the same point in time. It is also possible to combine cross section and time

series data (a *pooled* sample) to examine how demand varied across locations and over time. Each data source raises different estimation problems for analysts and requires different econometric methods.

Since economic theory specifies relationships between variables holding all else constant (the CETERIS PARIBUS assumption), researchers must be careful to ensure that the LINEAR RE GRESSION is properly specified. For example, when using time series data, changes in INFLA TION are likely, so variables must be adjusted to reflect real values. This problem is not likely to be significant in CROSS SECTION ANALYSIS.

See also *consumer price index; real income; real prices*

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Indus trial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Greene, W. H. (2002). *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.

Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.

time-series forecasting models

Roger Tutterow

Time series forecasting models are techniques used to explain changes in economic variables over time. As with other statistical methods, an underlying assumption is that the variable being modeled is stochastic, i.e., it contains a random component. Since data observations are realized over time, time series models differ from other statistical techniques in that the order in which the data are observed is of primary importance.

Time series models are classified as time domain or frequency domain models. In a fre quency domain model, the time series is decom posed into a sum of its cyclical components. This method, also known as spectral analysis, is beyond the scope of this article. For a discussion of spectral analysis, see Granger and Newbold (1986) or Hamilton (1994).

If a single variable is to be explained in terms of its own past values, then the model is said to

be univariate. For example, a model which fore casts an interest rate by using only the past values of the interest rate is a univariate model. In contrast, multivariate models consider how a set of variables interacts over time. A model which explains the relationship between the interest rate, money supply, and GROSS DOMESTIC PRODUCT (GDP) is a multivariate model.

An important requirement for most time series models is that the variable under consider ation follows a stationary process. A process is said to be weakly stationary if its mean and variance are constant over time, i.e., they exhibit no trend. If the process is also normally distrib uted, then it meets a stricter stationarity condi tion. While many economic and financial variables are not stationary, one can usually transform them into variables which are station ary. In subsequent discussions, we assume that the variable being forecast is either stationary or has been transformed into a stationary process.

The most basic methods for forecasting TIME SERIES DATA are moving average and exponential smoothing models. These tech niques range from simple models, in which the forecast value for a given period is the average of its past values, to more sophisticated models that contain error correcting components. The latter models include various exponential smoothing methods in which forecasts are corrected for past forecasting errors.

A second approach to forecasting is time series decomposition. In this technique, vari ations in the data are separated into trend, sea sonal, cyclical, and random components. The interaction of these components may be either additive or multiplicative. A time series is de composed through a series of iterations in which the individual components are extracted from the overall series. For example, by calculating a moving average of length longer than the sea sonal component, one may separate the trend and cyclical components from the seasonal and random components. Similar procedures allow the remaining individual components to be identified. Finally, the systematic components (trend, cyclical, and seasonal) are combined to generate a forecast.

The selection of an appropriate forecasting technique requires consideration of data avail

ability as well as time and resource constraints. Although the methods described above may be outperformed by more sophisticated models, they remain popular among managers due to their intuitive appeal and ease of use. Makrida kis, Wheelwright, and McGee (1983) provide an excellent critical survey of these techniques.

Perhaps the most important class of forecast ing models are autoregressive moving average (ARMA) models. In ARMA models, the forecast for a variable is based upon the past values of that variable, past forecasting errors, and a contem poraneous error term. By autoregressive, we mean that the current value of the variable depends upon its own past observations. For example, if the current INFLATION rate depends upon the "p" most recent inflation rates, then one says that inflation follows an autoregressive model of order p. Letting x_t denote the inflation rate at time t, this model is written as:

$$x_t = \Phi_1 x_{t\ 1} + \ldots + \Phi_p x_{t\ p} + u_t \qquad (1)$$

where Φ_i denotes the autoregressive parameters.

Alternatively, a variable may be generated as the weighted average of previous random dis turbances. If the variable is generated by the "q" most recent disturbances, then we say that it follows a moving average process of order q and may be written as:

$$x_t = u_t - \theta_1 u_{t\ 1} - \ldots - \theta_q u_{t\ q} \qquad (2)$$

where $u_{t\ 1}, \ldots, u_{t\ q}$ are serially independent random terms (the error terms have no AUTO CORRELATION) and the θ_i are the moving average coefficients.

Time series variables may contain both auto regressive and moving average components. For example, a variable may follow a process which is pth order autoregressive and qth order moving average. This model, denoted ARMA(p,q), is written as

$$x_t = \Phi_1 x_{t\ 1} + \ldots + \Phi_p x_{t\ p}$$
$$+ u_t - \theta_1 u_{t\ 1} - \ldots - \theta_q u_{t\ q}$$

It is common to represent ARMA models with "backshift" or "lag" notation. Define B as the operator which lags (moves) an observation back

one period such that $Bx_t = x_{t-1}$. For example, if x_t is the unemployment rate for the current period, then Bx_t denotes the unemployment rate for last period. In general, $B^d x_t$ denotes the unemployment rate d periods ago. Using back shift notation, the ARMA(p,q) model may be rewritten as

$$x_t(1 - \Phi_1 B - \ldots - \Phi_p B^p)$$
$$= u_t(1 - \theta_1 B - \ldots - \theta_q B^q)$$

As previously noted, most times series models, including ARMA models, require that the underlying process is stationary. While many economic variables are not stationary, they often may be transformed into a stationary process. The most common such transformation is differencing. Let x_t denote the observation on variable x at time t. Then, $x_t - x_{t-1}$ or $x_t - Bx_t$ is then called the first difference of the variable. Differencing is a particularly useful transformation for time series which include a time trend and thus are nonstationary. In this case, a first difference is frequently sufficient to remove the trend. In other cases, it may take repeated differencing to obtain a stationary process. Formally, if one must difference a series "d times" to render it stationary, then it is written as $x_t(1 - B)^d$ and the series is said to be integrated of order d. Thus, the ARMA(p,q) model generalizes to an ARIMA(p,d,q) where p, d, and q refer to the order of the autoregressive component, the degree of differencing required to make the process stationary, and the order of the moving average process, respectively. The ARIMA(p,d,q) model is written as

$$x_t(1 - B)^d(1 - \Phi_1 B - \ldots - \Phi_p B^p)$$
$$= u_t(1 - \theta_1 B - \ldots - \theta_q B^q)$$

Perhaps the most challenging aspect of using ARIMA models is the specification of the model, i.e., the selection of autoregressive, integration, and moving average orders. The most widely accepted methodology for model specification is that suggested by Box and Jenkins (1970).

Following Box and Jenkins, models are developed in three steps: identification, estimation, and diagnostic testing. During the identification phase, one inspects the plot of correlation coefficients between data observations and their lagged values. Specifically, one calculates both the simple and partial correlation coefficients between x_t and x_{t-1}, x_t and x_{t-2}, x_t and x_{t-3}, etc. By inspecting the plots of these coefficients, also known as correlograms, one may infer the likely structure of the model. Specifying a model by inspecting the correlograms is not a mechanical process but rather requires experience.

After a model structure is selected, the parameters of the model, $\Phi_1, \ldots, \Phi_p, \theta_1, \ldots, \theta_q$, are estimated simultaneously by the maximum likelihood method. Finally, the residuals of the model are subject to diagnostic tests. If these tests are passed, the residuals are considered "white noise" (random) and the model is deemed suitable for forecasting use. Failure of these tests suggests that the model is incorrectly specified and must be revised.

In some instances, it may be of interest to generate forecasting models which utilize information contained in other economic variables. For example, a manager may wish to incorporate information on ADVERTISING expenditures into a sales forecast. This is a simple example of a multivariate time series model. One should note that for multivariate time series, the concept of a stationary process is complicated by the possibility of cointegration, i.e., linear combinations of variables may be stationary even if the individual variables are not.

A popular multivariate extension of the univariate ARIMA model is the transfer function model. This model extends the Box–Jenkins methodology to allow changes in an exogenous "input" variable to induce changes in an endogenous "output" variable. By inspecting a plot of the correlation coefficients between the output variable and previous values of the input variable, one may infer the temporal relationship between the variables. For example, one might find that the inflation rate is highly correlated with changes in the money supply which occurred three months prior. To assist with identification, it is common to filter or "pre whiten" both variables with an ARIMA model before estimating the correlation coefficients. Having identified the temporal relationship between the variables, the transfer function is estimated.

Finally, an ARIMA model is fit to the residuals of the transfer function and diagnostic tests are performed.

Another useful multivariate time series model is vector autoregression (VAR). This technique, popularized by Sims (1980), ex plains changes in a vector (or set) of variables in terms of past observations on those variables. Unlike transfer function models, VAR does not impose restrictions upon the exogeneity of a specific variable, but rather, it allows changes in variables to "feed back" into the other vari ables in the model. Advocates of VAR argue that the technique is useful since it imposes few a priori restrictions upon the structure of the model. One disadvantage of VAR models is that they sometimes require estimation of a large number of parameters. In practical application, if data sets are small, the parameter require ments of a VAR model may be prohibitive. While most early applications of VAR were to macroeconomic models, applications in micro economics and financial economics are becom ing more common.

Recent advances in time series forecasting models include tests for cointegration of time series, analysis of models in which the variance of the error term varies over time (ARCH), and the development of unit root test. These topics are beyond the scope of this article. Interested readers are referred to Enders (1995) and the citations contained therein.

Bibliography

Box, G. and Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Hol-den-Day.

Enders, W. (1995). *Applied Econometric Time Series*. New York: John Wiley.

Granger, C. and Newbold, P. (1986). *Forecasting Eco nomic Time Series*. San Diego: Academic Press.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Makridakis, S., Wheelwright, S., and McGee, V. (1983). *Forecasting: Methods and Applications*, 2nd edn. New York: John Wiley.

Sims, C. (1980). Macroeconomics and reality. *Econome trica*, **48**, 1 49.

Wilson, J. H. and Keating, B. (2002). *Business Forecasting*, 4th edn. New York: McGraw-Hill.

total costs

Gilbert Becker

The total costs facing any business are the sum of two distinct categories of costs:

1   TOTAL VARIABLE COSTS and
2   total FIXED COSTS.

The total cost function is the relation between the total dollar expenditures for resources by a firm plus OPPORTUNITY COSTS and the level of output and sales achieved by the firm. The total costs of a firm naturally tend to increase with the level of output and sales effort as more resources are needed to achieve this growth. In the SHORT RUN the total costs eventually tend to increase at an increasing rate due to the de clining marginal productivity of additional units of variable inputs being combined with fixed inputs (*see* DIMINISHING RETURNS). In the long run, where the size of fixed inputs can be adjusted, the total cost of production may ini tially increase at a decreasing rate as output grows due to ECONOMIES OF SCALE. Ordinar ily, at some higher rate of output, total costs will tend to increase at a constant rate and will ultim ately increase at an increasing rate (*see* MIN IMUM EFFICIENT SCALE).

The total costs of a firm are important in determining the firm's profit. Since profit is measured as TOTAL REVENUE minus total costs, efforts that reduce total costs without de creasing revenues will enhance profits. Econo mists distinguish between ECONOMIC PROFIT and ACCOUNTING PROFIT. This distinction is based entirely on the differing methods which the two disciplines employ in measuring costs. The difference lies partially in the inclusion of opportunity costs in the measure of total eco nomic costs.

See also *long run cost curves; short run cost curves*

Bibliography

Maurice, S. and Thomas, C. (2002). *Managerial Econom ics*, 7th edn. Boston: McGraw-Hill Irwin.

Nicholson, W. (2000). *Intermediate Microeconomics and its Application*, 8th edn. New York: Harcourt.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

total product

Robert E. McAuliffe

Total product refers to the output produced by varying the level of one input, holding the level of other inputs constant. Thus the total product of labor is the total output produced by changing labor inputs while the other factors of production are not changed (including the level of technology). The total product curve can be derived from the production function in the SHORT RUN (*see* PRODUCTION FUNCTIONS). The typical total product curve is S shaped where the first units of labor, for example, are very productive and output increases rapidly. But as more workers are added to the fixed amounts of capital, diminishing marginal returns occur and the increases in output from additional workers diminish. The total product curve is the foundation for the firm's SHORT RUN COST CURVES.

See also *DIMINISHING RETURNS; MARGINAL COST; MARGINAL PRODUCT*

Bibliography

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

Shughart, W. F., Chappell, W. F., and Cottle, R. L. (1994). *Modern Managerial Economics*. Cincinnati, OH: South-Western Publishing.

total revenue

Gilbert Becker

Total revenue is the total amount of dollars received by a firm or an industry from the sale of its product in a given time period. Often known as sales revenue or sales receipts, its measure is equivalent in value to the total dollar expenditures that are made by the consumers of a given product. These revenues are determined by the price (*P*) of the product and the quantity (*Q*) demanded at that price, and are calculated as:

$$TR = P \times Q$$

The quantity demanded depends upon consumer tastes, income, and other factors (*see* DEMAND FUNCTION), while price is determined by market supply forces as well.

Under PERFECT COMPETITION, total revenue increases in proportion with unit sales by the firm since price is fixed. For imperfectly competitive markets, e.g., OLIGOPOLY, firms maintain some control over the price of their product. As such, managers must be aware of the LAW OF DEMAND, which indicates that any price change for a good will typically induce an opposite change in its quantity demanded, and thus the change in the total revenue of the firm is uncertain. Here, as consumer demand for a good typically declines with an increase in the price of the good, the size of the consumer reaction is critical to the overall impact on total revenue (*see* ELASTICITY). Alternative strategies, such as increasing ADVERTISING or improving product QUALITY, may raise unit sales without requiring a reduction in price and thus would increase total revenue.

Total revenue is also important in the calculation of the firm's profits, since profit is measured as total revenue minus TOTAL COSTS. Maximizing total revenue is sometimes a goal of managers, although this goal by itself is not equivalent to that of PROFIT MAXIMIZATION as it ignores costs and the level of profits. For a product with a given demand curve, revenue maximization requires choosing an output level where MARGINAL REVENUE is zero (*see* DEMAND CURVES). For a firm selling to two or more well defined sets of consumers, maximizing total revenue requires that the marginal revenue from each group be equal (*see* PRICE DISCRIMINATION).

See also *inelastic demand*

Bibliography

Browning, E. and Zupan, M. (2004). *Microeconomics: Theory and Applications*, 8th edn. New York: John Wiley.

Maurice, S. and Thomas, C. (2002). *Managerial Economics*, 7th edn. Boston: McGraw-Hill Irwin.

total utility

Gilbert Becker

Total utility is the total benefit or satisfaction an individual receives by consuming goods or taking part in an activity. For example, consumers acquire utility from the product attributes in the goods and services that they purchase (*see* PRODUCT ATTRIBUTES MODEL). These items fulfill various physical and psychic needs. Similarly, workers gain utility from their jobs, from the satisfaction of accomplishment, and also from the income that is generated. While consumer utility arises from the goods themselves, giving to others (e.g., gifts, charity) also generates utility, as can the actual purchase of the good (e.g., "finding a real bargain"). An individual's preferences toward risk can also be identified, through an examination of the level of utility generated from different levels of income (*see* RISK AVERSION).

In the economic theory of consumer behavior, total utility is typically assumed to increase with an increase in the number of units of any one good, and with an increase in the total number of all goods. In other words, for most items, known as **economic goods**, more is preferred to less. As the rate of consumption per period of any one good grows, however, the total utility tends to increase at a decreasing rate, as the first units of the good provide the highest utility and additional units eventually generate less utility (*see* DIMINISHING MARGINAL UTILITY).

Two exceptions to the assumption that more is preferred to less are worth noting. The first is a group of items, known as **economic bads**, which create negative marginal utility – i.e., a decline in total utility – as more is consumed.

For example, more jail time is an economic bad for most criminals, as is greater risk for many investors. The second group, known as **neutral goods**, are items which provide no utility (or disutility) at all to the consumer. Ashtrays, for example, may be considered to be neutral goods by many nonsmokers, who see this item as essentially useless.

Although some models of economic behavior examine the situation where the individual's goal is to achieve a satisfactory level of utility, it is generally assumed that UTILITY MAXIMIZATION is the objective. Here, INDIFFERENCE CURVES are useful. This analytical tool measures utility ordinally, and thus only requires that the individual be able to rank any available choice as yielding more, less, or the same amount of total utility as other alternatives. Earlier consumer models, which measured utility cardinally, were weaker in that they assumed that the exact amount by which MARGINAL UTILITY differs across choices could be specified. Identifying, and perhaps altering (e.g., through ADVERTISING), the underlying sources of consumer utility is central to understanding and generating consumer demand for a product.

Finally, total utility is important in the identification and measurement of CONSUMER SURPLUS and allocative efficiency (*see* EFFICIENCY). Consumer surplus is measured as the difference between the total utility received from the purchase of a good and the total amount paid for that good by consumers. Allocative efficiency in a market requires that the sum of consumer surplus plus PRODUCER SURPLUS be maximized.

See also *deadweight loss*

Bibliography

Browning, E. and Zupan, M. (2004). *Microeconomics: Theory and Applications*, 8th edn. New York: John Wiley.

Maurice, S. and Thomas, C. (2002). *Managerial Economics*, 7th edn. Boston: McGraw-Hill Irwin.

Nicholson, W. (2000). *Intermediate Microeconomics and its Application*, 8th edn. New York: Harcourt.

Varian, H. R. (1992). *Microeconomic Analysis*, 3rd edn. New York: Norton.

total variable cost

Lidija Polutnik

Total variable cost measures the total variable costs of production in the short run. Total variable costs are those costs which vary with the output level produced, such as workers' wages, material inputs, etc. Total variable cost depends on the quantity of output produced and is equal to zero when no output is produced. In addition to the per unit input cost, the MARGINAL PRODUCT of inputs and the technology employed by the company influence the level of total variable cost (*see* SHORT RUN COST CURVES).

The curvature of a total variable cost curve is determined by the impact of successive increases in output on total variable cost. Typically, total variable costs will increase initially at a decreasing rate but will eventually increase at an increasing rate because of DIMINISHING RETURNS in production (*see* AVERAGE VARIABLE COSTS; TOTAL COSTS; and the entry on marginal product).

Labor costs are usually considered as variable in nature, i.e., they can be reduced if the worker works fewer hours and vice versa. However, there are many examples of labor costs which are not strictly related to the number of hours of work and can be defined as quasi fixed. Examples are employee benefits such as health insurance, pension plans, costs of hiring and training new employees, and the costs of legally required social insurance programs. The company pays for these costs on a per worker rather than per labor hour worked basis. A company's decision about the optimal combination of inputs in its production process will be influenced by the amount of these quasi fixed costs per worker. For example, part time employees are less likely to be covered by the company's health insurance policy and are usually given fewer fringe benefits in general, making them a relatively cheaper input in production than a full time worker.

Bibliography

Carlton, D. W. and Perloff, J. M. (2000). *Modern Industrial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

transactions costs

Lidija Polutnik

Transactions costs are the expenses in time and resources that are associated with the process of buying and selling other than the price. The level of these costs depends on the type of transaction conducted as well as its organizational form. According to Coase (1937), a firm emerges as an organization of economic activity when it can economize on the transactions costs by performing activities within its boundaries rather than having these activities performed by the market. Some examples of transactions costs when the firm uses the market are: searching for the best price, best quality and delivery terms, the costs of writing a contract, monitoring and enforcing the agreement as well as bargaining costs. If the firm decides to conduct activities itself, it faces a different set of transactions costs. Some examples of these are the costs of coordinating purchases and production activities and the costs of motivating and monitoring the work force to ensure best behavior for the firm.

According to Williamson (1979), the advantages of organizing activities within a firm and for firms to integrate (*see* VERTICAL INTEGRATION) are greater when transactions have certain characteristics. ASSET SPECIFICITY creates incentives for opportunistic behavior where a contract or practice is needed to protect the participant against premature renegotiation or cancellation of a contract. Vertical integration may also enable a firm to economize on transactions costs when the degree of COMPETITION in supplying industries is low and there is a possibility of a supplier holding up production. Managers may not be able to enumerate and articulate all the contingencies of a contract due to bounded rationality which present possibil

ities for opportunistic behavior as well (*see* IN-COMPLETE CONTRACTS). When transactions occur frequently and over a long period of time, it may be more efficient to conduct these trans-actions within a firm. Transactions costs in-crease as the complexity of a transaction increases and as the predictability of the market for an activity decreases. The higher the trans-actions costs, the more likely the firm will choose to organize that activity within its organization.

Transactions cost economics can be employed to analyze the structure, conduct, and performance across industries (*see* STRUC-TURE CONDUCT PERFORMANCE PARA-DIGM); it enables us to evaluate contractual relations among and within firms as a result of organizations economizing on transactions costs when information is incomplete and enforce-ment problems exist.

See also *Coase theorem; principal–agent prob-lem*

Bibliography

Brickley, J. A., Smith, C. W., and Zimmerman, J. L. (2000). *Managerial Economics and Organizational Archi-tecture*. Chicago: McGraw-Hill.

Carlton, D. W. and Perloff, J. M. (2000). *Modern Indus-trial Organization*, 3rd edn. Reading, MA: Addison-Wesley.

Coase, R. H. (1937). The nature of the firm. *Economica*, **4**, 386 405.

Milgrom, P. and Roberts, J. (1992). *Economics, Organiza-tion and Management*. Englewood Cliffs, NJ: Prentice-Hall.

Williamson, O. (1979). Transaction-cost economics: The governance of contractual relations. *Journal of Law and Economics*, **22**, 233 61.

transfer pricing

Dileep R. Mehta and James G. Tompkins

A transfer price is an internal price that is charged for a product or service by a selling division to a buying division within the same corporation (see Hirshleifer, 1956, 1957; Copithorne, 1971; Horst, 1971, 1977). The prime goal of a transfer pricing mechanism is to maximize global firm value. Specifically,

transfer pricing can affect firm value in two prime areas: tax minimization and optimal re-source allocation.

Minimizing global taxes can be achieved when a company with operations in countries with different tax bases shifts profits from high to low tax areas. One way to do this is to charge a high transfer price by a producing division in a low tax country to a consuming division in a high tax country. Similar strategies can also be implemented to minimize tariff charges and the effects of exchange controls. However, since a country's tax revenue is affected by a company's transfer pricing mechanisms, international tax treaties and local laws constrain the methods by which a transfer price for tax purposes may be set.

Transfer pricing also affects how the divisions within a firm allocate their resources. Since firm value is maximized when its limited resources are allocated optimally, by implication transfer pricing affects firm value. For example, if at one extreme a firm were to charge an excessive price for a facility, potential users would shun the facility and it would have EXCESS CAPACITY. At the other extreme, if its use is permitted free of charge, there is little incentive for the con-sumers to use it in a cost effective manner. Thus an appropriate transfer price would avoid excess capacity on the one hand, and would not precipi-tate additional value decreasing investment outlays on the other.

Transfer pricing affects optimal resource al-location when a company has two or more decentralized divisions whose profit or cost per-formance is independently measured and their incentive schemes are tied to these per-formance measures. To maximize firm value, the transfer price between the two divisions should be the opportunity cost of the good or service transferred (*see* OPPORTUNITY COSTS). The opportunity cost of the transferred unit is defined as the value the producing div-ision could have received if the unit had been employed in its next best alternative use. Unfor-tunately, an accurate opportunity cost is difficult to ascertain. Partly this is because division man-agers may jealously guard information necessary to calculate an opportunity cost which has a direct impact on performance evaluation by their superiors. In addition, even in the presence

of perfect information, cost classification issues make an opportunity cost very difficult or costly to assess. Nevertheless, three common methods of transfer pricing have emerged: market based, cost based, and negotiated price.

In a market based approach, the opportunity cost is proxied by the price that prevails for a like good or service in a competitive open market (*see* PERFECT COMPETITION). An advantage of this system is that the price is externally observable and price distortion between the two divisions is minimized. This method is also favored by tax authorities since they are able to use information that is external to the firm to determine whether they have received fair tax revenues. A disadvantage is that a market price for a similar good in a competitive market is not always available. Furthermore, even if an external market does exist for the intermediate good, the assumption of a competitive market is likely to be erroneous. This is because the preference for internal production implies BARRIERS TO ENTRY and/or EXIT which render the assumption of a competitive market fallacious. This is highlighted by ECONOMIES OF SCALE and/or ECONOMIES OF SCOPE as well as QUALITY control standards or other SYNERGIES which may exist when the intermediate good is produced internally. In the final analysis, once the unrealistic premise of competitive markets is removed, PREDATORY PRICING practices and their like cannot be ruled out and an exclusive focus on market price is unlikely to be the optimal transfer price for the firm (see Deschamps and Mehta, 1989).

In the absence of an external market for the intermediate good, a company will often resort to setting a transfer price based on costs. In principle, the MARGINAL COST of the transferred good may closely represent its opportunity cost. However, the selling division has an incentive to exaggerate its classification of costs as marginal. Another problem is that even if marginal costs are fairly classified, the selling division may not show a profit since no contribution toward FIXED COSTS is incorporated in the price. Some companies resolve this problem by pricing at marginal cost plus a premium. At the extreme, many companies permit a transfer price at full cost representing the sum of marginal and fixed costs (see Brickley, Smith, and Zimmerman, 1995). Such a system, however, removes the incentive by the selling division to provide the good or service in the most cost effective manner.

A severe disadvantage inherent in any cost based transfer price method is that most firms rely on accounting data. This approach is erroneous since identifying the opportunity cost and hence the transfer price is an economic as opposed to accounting issue. For example, large differences in costs can emerge when fixed costs are depreciated over an accounting life as opposed to an economic life (*see* ECONOMIC DEPRECIATION). In practice, however, because of the availability, many firms resort to accounting data when establishing a transfer price.

One solution to the conflict of interest that exists between a producing and consuming division is to have a system of negotiation when setting transfer prices. In such a system, the selling and buying divisions negotiate both a price and quantity of transferred goods or services. The problem with this is that the transfer price may not represent the opportunity cost and hence firm value is not maximized. Part of the reason is that the transfer price depends on the negotiation skills of the two divisions. Another disadvantage to price negotiation is that it can entail significant time and effort on the part of the negotiators.

In summary, transfer pricing affects firm value in two major ways. The first is the role it plays in taking advantage of global market imperfections such as differences in tax structures and other regulations among countries. The second is the effect it has on the optimal allocation of a firm's resources. In theory, a firm's resources are optimally allocated when the transfer price is set at the opportunity cost of the good or service transferred. If an external competitive market exists for the intermediate good, then the opportunity cost can be proxied at this external market price. However, an external market price does not necessarily represent opportunity cost as it does not incorporate internal synergies that may exist with internal transfers. These differences and the preference for an internal market make the assumption of a competitive market structure fallacious. In this case, firms may resort to cost or negotiation methods of transfer pricing which also have their weaknesses. Irre

spective of the methods discussed, there ultimately exists an inherent conflict of interest within decentralized organizations between setting a transfer price that maximizes division profitability and a transfer price that maximizes total firm value.

Bibliography

Brickley, J., Smith, C., and Zimmerman, J. (1995). Transfer pricing and the control of internal corporate transactions. *Journal of Applied Corporate Finance*, **8**, 60 7.

Copithorne, L. W. (1971). International corporate transfer prices and government policy. *Canadian Journal of Economics*, **4**, 324 41.

Deschamps, B. and Mehta, D. (1989). *Chief Financial Officer: Strategy Formulation and Implementation*. New York: John Wiley.

Hirshleifer, J. (1956). On the economics of transfer pricing. *Journal of Business*, **29**, 172 84.

Hirshleifer, J. (1957). Economics of the divisionalized firm. *Journal of Business*, **30**, 96 108.

Horst, T. (1971). Theory of the multinational firm: Optimal behavior under differing tariff and tax rates. *Journal of Political Economy*, **79**, 1059 72.

Horst, T. (1977). American taxation of multinational firms. *American Economic Review*, **67**, 376 89.

translog cost function

Roger Tutterow

When investigating properties of a production process, it is common to represent the process with a production function (*see* PRODUCTION FUNCTIONS). Shephard (1953) has shown that, under certain regularity conditions, a cost function provides an economically equivalent representation of a production function. Thus, while some econometric studies have estimated the production function directly, there are frequently advantages to estimating a cost function instead.

In estimating these cost functions, the TOTAL COSTS of production are modeled as depending on the level of output produced and the prices of the factors of production that are used in the process. In the cases of firms that produce multiple products, the total cost will depend on the levels of output of each product. In either case, estimating the cost function requires one to assume a functional form for the relationship between the levels of the outputs, the prices

of the factors of production, and the total cost of production. The choice of functional form imposes some restrictions on the properties of the underlying production process that may not be defensible. An alternative is to estimate a cost function using a "flexible functional form."

One such flexible functional form is the translog (or transcendental logarithmic) form. Drawing on the work of Christensen, Jorgenson, and Lau (1975), the translog cost function for a multiproduct firm is given by

$$
\begin{aligned}
\ln C = \alpha_0 &+ \sum_{i=1}^{m} \alpha_i \ln Y_i \\
&+ \sum_{j=1}^{n} \beta_j \ln W_j \\
&+ \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \delta_{ij} \ln Y_i \ln Y_j \\
&+ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} \ln W_i \ln W_j \\
&+ \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} \rho_{ij} \ln Y_i \ln W_j
\end{aligned}
$$

where C is the total cost of production, Y_i is the level of output of the ith product, and W_j is the price of the jth factor of production. If the production process under consideration has only one output, then the cost function simplifies accordingly. The translog specification is "flexible" in the sense that it provides a "local approximation" to any cost function.

When estimating this function, one should impose some restrictions on the parameters of the model. First, by adding the restrictions that

$$
\sum_{j=1}^{n} \beta_j = 1 \quad \text{and that} \quad \text{for each} \quad \text{"i"}
$$

$$
\sum_{i=1}^{m} \delta_{ij} = 0, \sum_{j=1}^{n} \rho_{ij} = 0, \text{ we insure that the cost}
$$

function is linearly homogeneous in FACTOR PRICES. Second, symmetry conditions require that $\delta_{ij} = \delta_{ji}$ and $\rho_{ij} = \rho_{ji}$. These restrictions both insure that the estimated model has properties required of well behaved cost functions and improve the precision of the parameter estimates.

While one may estimate the translog cost function directly, the parameters can be estimated with greater econometric efficiency by estimating a system of cost share equations. Applying Shephard lemma, one may transform the translog cost function into a system of cost share equations that are then estimated using the method of seemingly unrelated equations developed by Zellner (1962).

After estimating the cost function, it is possible to test for ECONOMIES OF SCALE by testing restrictions on the coefficients. Likewise, for a multiple output production process, the model may also be tested for ECONOMIES OF SCOPE. The use of the multiproduct translog model to test for economies of scale and scope has been particularly popular in the telecommunications, energy, and financial services industries. For a more detailed treatment of single product and multiproduct cost functions, see Fuss and McFadden (1978) or Chambers (1988) and the references contained therein.

Bibliography

Chambers, R. (1988). *Applied Production Analysis*. Cambridge: Cambridge University Press.

Christensen, L. R., Jorgenson, D. W., and Lau, L. J. (1975). Transcendental logarithmic utility functions. *American Economic Review*, **65**, 367 83.

Fuss, M. and McFadden, D. (1978). *Production Economics: A Dual Approach to Theory and Applications*. Amsterdam: North-Holland.

Shephard, R. (1953). *Cost and Production Functions*. Princeton, NJ: Princeton University Press.

Zellner, A. (1962). An efficient method for estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, **58**, 977 92.

tying

Gilbert Becker

Tying occurs when the sale or lease of one good is made conditional on the purchase of a second good (or service). The first good is identified as the tying good while the second good is the tied good. For example, in the mid 1990s Microsoft Corporation, manufacturer of the leading personal computer operating system known as Windows, insisted that computer manufacturers wishing to purchase Windows also install Microsoft's Internet Explorer (IE) web browser (*see* MICROSOFT ANTITRUST CASE). This created a tie between the web browser (the tied good) and the operating system (the tying good).

These types of arrangements, which come in many forms, have the potential to cause economic harm to COMPETITION. As such they may violate Section 3 of the CLAYTON ACT, or, in cases of ties between goods and services (and other special cases), Section 1 of the SHERMAN ACT. The possible adverse effects from tying include:

1 the extension of MARKET POWER from the market for the tying good into the market for the tied good;
2 the reduction of freedom of choice by consumers; and
3 the foreclosure of rivals from access to markets.

This last consequence is often considered to be the most serious concern in antitrust cases.

Tying CONTRACTS have also been shown to offer economic benefits in some circumstances. These include:

1 production and distribution ECONOMIES OF SCALE; and
2 the protection of product QUALITY, e.g., by tying a service contract to the sale of a product.

Both of these benefits offer potential advantages to consumers and producers. Consumers stand to gain from lower prices due to the lower costs if the market is competitive. They also may benefit from the reduction of SEARCH COSTS. Included here is the time saving ease of purchase through one stop shopping afforded to customers. They also gain from their greater certainty of the product's performance. This may be especially true in markets involving complex high technology goods, where uncertainty concerning product compatibility between the two goods may otherwise exist.

Producers may gain greater profits from lower costs and also gain in their greater ability to protect their reputation. More importantly,

managers should be aware that tying may allow for additional profits for the firm, even where competition exists in the market for the tied good. If a firm has market power in the tying good market (e.g., copying machines), then by lowering the price of this good (and thereby inducing more purchases; *see* LAW OF DEMAND) the firm can charge a higher than competitive price in the market for the tied good (e.g., copy paper). Here the tie will often allow the producer to capture additional CONSUMER SURPLUS from high use consumers who have a more INELASTIC DEMAND. As such, tying here is considered a form of PRICE DISCRIMINATION.

Since tying may offer both economic benefits and costs, it is only illegal in the US where it "substantially lessens competition," and where the existence of some market power in the tying good's market has been demonstrated. In the European Union (EU), the legality of tying lies primarily in the question of whether the tie constitutes an "abuse of dominant position" by a leading firm in a market. In its own case against the Microsoft Corporation, the EU is currently reexamining its legal standards concerning tying (*see* EU COMPETITION POLICY, 2004).

Bibliography

Breit, W. and Elzinga, K. (1996). *The Antitrust Casebook: Milestones in Economic Regulation*, 3rd edn. Fort Worth, TX: Dryden Press.

Church, J. and Ware, R. (2000). *Industrial Organization: A Strategic Approach*. Boston: McGraw-Hill Irwin.

Fisher, F. and Rubinfeld, D. (2001). *US* v. *Microsoft*: An economic analysis. In W. Commanor (ed.), Symposium: Economists examine Microsoft its concepts, findings, and remedies. *Antitrust Bulletin*, **46**, 1 72.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

U

uncertainty

Robert E. McAuliffe

Frank Knight (1971) emphasized the distinction between risk and uncertainty. He defined risk as measurable or quantifiable, such as when a life insurance firm calculates the probabilities of paying benefits to survivors. The term risk applies to known probabilities and outcomes. However, in many business situations managers cannot reasonably calculate probabilities and may not know the possible outcomes in the future. In these cases where risk cannot be quantified, Knight argued there is uncertainty. Situations involving uncertainty as defined by Knight are much more difficult for decision making and this is why profits are awarded to business people: as a reward and a return for significant risk bearing.

Bibliography

Douglas, E. J. (1992). *Managerial Economics*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Knight, F. H. (1971). *Risk, Uncertainty and Profit*. Chicago: University of Chicago Press.

utility maximization

Gilbert Becker

In most economic models it is assumed that individuals attempt to maximize their TOTAL UTILITY. Ordinarily, these individuals must choose between alternatives and are faced with constraints. For example, consumers purchase a set of goods or product attributes which give them the most satisfaction given their limited income (*see* PRODUCT ATTRIBUTES MODEL). Workers select a number of hours to work based on their preferences between work and leisure, subject to a time constraint and market constraints including the wage rate being offered. Managers and owners maximize their satisfaction by making investment decisions based on their preferences toward risk and rates of return, often subject to a constraint of limited funds. These and other decisions can be examined using INDIFFERENCE CURVES.

The model of consumer behavior typically describes the choice between two goods, for example, X and Y, or between one good and a composite good (the remaining funds for all other goods). The indifference curves of any individual show that consumer's tastes toward these two items. The slope of these indifference curves shows the rate at which the consumer is willing to trade one good for another, which depends on the ratio of the MARGINAL UTILITY (MU) of one good, X, to that of the other good, Y (*see* MARGINAL RATE OF SUBSTITUTION). The consumer is limited by a BUDGET CONSTRAINT which identifies the set of choices actually available given her income and the prices of the goods. The consumer's goal is to choose from this obtainable set that bundle which gives the most satisfaction.

The condition for utility maximization can be formally written as:

$$\frac{MU_x}{MU_y} = \frac{P_x}{P_y} \qquad (1)$$

where the ratio on the left hand side of the equation shows the rate at which the consumer is willing to trade the two goods. The right hand side of the equation measures the ratio of the

prices (P) of the two goods, which indicates the market rate at which the consumer must trade the two goods. For example, assume that the consumer was willing to trade three units of good Y for one extra unit of X (since her MU_x was three times that of Y). Here, the ratio of the marginal utilities would be 3 to 1. If the market price of X was only twice that of Y, the consumer would only have to give up two units of Y to obtain the additional unit of X desired. Greater utility would be achieved by making this trade and by continuing to sacrifice more units of Y for additional X until the equality as stated in equation (1) is achieved.

Formally, this is described in figure 1. The consumer chooses a combination of goods Y and X whose quantities (Q) are measured on the vertical and horizontal axes. The indifference curves are marked i_0, i_1, $i_2 \ldots$, with higher numbers indicating greater total utility. The budget line, typically linear, has a slope that is determined by the ratio of the market determined prices. Point E represents utility maximization for the consumer. This is the tangency point between the budget line and the highest indifference curve within the budget constraint. Points on higher indifference curves, while desirable, are unachievable given current prices and the consumer's income. Points below the budget line, while obtainable, offer less total utility than E as they fall on lower indifference curves. The tangency at E indicates that the slopes of the indifference curve and the budget line are equal, and as such the quantity of good Y which the consumer is willing to trade for additional units of good X is equal to the amount she has to trade. The combination of Q_Y^* and Q_X^* will be chosen by the utility maximizing consumer.

From figure 1, a strategy of lowering the price of X, P_x, can be shown to alter the budget line in such a way that a higher indifference curve can be achieved. The new utility maximization point typically holds a greater quantity of good X, which is consistent with the LAW OF DEMAND. Moreover, the size of the INCOME EFFECT and SUBSTITUTION EFFECT of the price change can be derived from this model. Alternatively, it can be demonstrated that a change in con
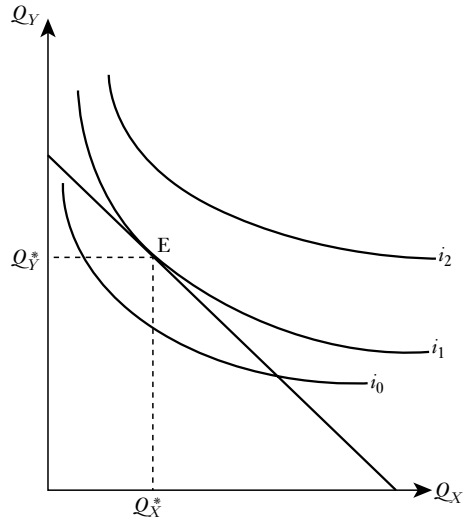
Figure 1   Utility maximization

sumer tastes, perhaps resulting from ADVERTISING or other promotion of good X, would increase the steepness of the slopes of all of the curves in the indifference map by altering the marginal rates of substitution. Once again, a new utility maximization point would be found. While remaining on the initial budget line, the consumer again will typically purchase more of good X.

One example of how managers may use this utility maximization model involves a marketing survey received by the author from one of the largest banks in the US. Fleet Bank was considering the introduction of two new checking account options and wished to identify consumer preferences and willingness to pay for these new services. The survey issued a set of cards indicating different bundles of options (*see* BUNDLING) that may be made available, and asked survey respondents to rank each card at one of five levels of utility according to their preferences. All cards placed in level 1, for example, created a low level indifference curve. The introduction of prices in some of the cards enabled the market researchers to statistically estimate which new products offered the most utility and also an optimal pricing strategy for the firm.

Numerous other applications of this model have been generated. Lancaster (1971) analyzes consumer demand for products by developing indifference curves for product attributes. Nicholson (2000) and others demonstrate the conditions under which food stamp programs are effective, ticket scalping and cost of living adjustments for recipients of government assistance are justifiable, and frequent flyer plans are profitable.

Bibliography

Lancaster, K. (1971). *Consumer Demand: A New Approach*. New York: Columbia University Press.

Maurice, C. and Thomas, C. (2002). *Managerial Economics*, 7th edn. Boston: McGraw-Hill Irwin.

Nicholson, W. (2000). *Intermediate Microeconomics and its Application*, 8th edn. New York: Harcourt.

Perloff, J. M. (2004). *Microeconomics*, 3rd edn. Boston: Pearson Addison-Wesley.

V

variance

Robert E. McAuliffe

The variance, or second moment, of a random variable measures the dispersion of outcomes around that variable's EXPECTED VALUE. For a discrete random variable, X, which has specific outcomes denoted by X_i and where the prob ability distribution of the values of X is given by $f(x_i)$, then the variance of X, σ^2, is calculated as

$$\text{var(X)} = \sigma^2 = \sum_x f(x_i) \times (X_i - \text{E}(X))^2 \quad (1)$$

where $\text{E}(X)$ is the expected value of the random variable, X, and $\sum_x$ denotes the summation over all possible values of X. Each squared deviation of the outcome, X_i, from the expected value, $\text{E}(X)$, is weighted by its probability, $f(x_i)$.

For a given sample with T observations, an unbiased estimator of the variance of a random variable is s^2 which is calculated as

$$s^2 = \frac{1}{T-1} \times \sum_i^T (X_i - \bar{X})^2 \quad (2)$$

where $\bar{X}$ is the sample mean. Note that the squared deviations are divided by $T - 1$ and not by T because sample information has already been used to calculate the sample mean, so only $T - 1$ of the observations are independent (see Maddala, 2001; Greene, 2002). Most numerical software programs and spreadsheets will calcu late the sample variance for a set of observations.

Normally, the greater the variance (or disper sion) of possible outcomes around the mean, the greater the risk to a manager or investor. The variance may thus be used to measure risk (or the square root of the variance, the STANDARD DE VIATION).

Bibliography

Greene, W. H. (2002). *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.

Gujarati, D. N. (2002). *Basic Econometrics*, 4th edn. New York: McGraw-Hill Irwin.

Keat, P. G. and Young, P. K. Y. (2002). *Managerial Economics: Economic Tools for Today's Decision Makers*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.

Maddala, G. S. (2001). *Introduction to Econometrics*, 3rd edn. New York: John Wiley.

vertical integration

Govind Hariharan and Sean Hopkins

Vertical integration refers to the expansion of a firm into earlier or later stages in the production process. For example, when General Motors decides to manufacture car radiators in its own plants instead of purchasing them from external sources, it is vertically integrating.

There are two types of integration. Forward ("upstream") integration occurs when a firm decides to produce a product or service that uses this firm's product or service as an input (e.g., GM starts selling cars at company owned dealerships). Backward ("downstream") integration (*see* BACKWARD INTEGRATION) occurs when one firm decides to produce the inputs itself rather than purchasing them from outside sources (e.g., GM producing radiators). The distinguishing feature of any vertical inte gration is the replacement of a market purchase by an internal transfer. This could occur as a result of the acquisition of other firms or as an expansion within a firm. Assuming that firms are organized for the purposes of earning profits, a primary reason for replacing a market exchange

with an internal transfer should be that, for that particular input, the latter mechanism is less expensive than the former.

Nobel economist Ronald Coase was interested in finding out why people organize production activities in the hierarchical structure of the firm (*see* HIERARCHY) and coordinate their decisions through a central authority rather than relying on market exchange (Coase, 1937; see COASE THEOREM). His explanation was that organizing activities through the hierarchy of a firm is often more efficient than market exchange because production requires the coordination of many transactions among many resource owners (*see* EFFICIENCY). According to Coase, the costs of transacting business through market relations are often higher than those of undertaking the same activities within the firm. Coase's major insight was that economic activity is most easily understood in terms of the TRANSACTIONS COSTS involved in any system of exchange between individuals.

Coase's analysis of transactions costs and vertical integration provides a good starting point for understanding the organizational decisions of the firm, but it falls somewhat short of a complete treatment in two important aspects (Schupack, 1977). First, Coase provides little detail concerning the underlying sources of transactions costs. This void limits the predictive use of the analysis. Moreover, Coase implicitly assumes that a competitive intermediate products market exists, so that MARKET STRUCTURE influences on vertical integration and their consequences simply do not arise.

Although Coase mentions SEARCH COSTS and the problems of long term CONTRACTS, the most extensive treatment of transactions cost determinants has been carried out by Williamson (1974). He identified two sets of factors whose elements interact to increase the costs of market exchange. The first set of factors, which Williamson referred to as "transactional factors" (market UNCERTAINTY and trading partners), is concerned with the environmental characteristics of the relevant intermediate product market. According to Williamson, market uncertainty increases transactions costs since a greater degree of uncertainty in a market will result in the negotiation of more lengthy and complex contracts between buyers and sellers of the inter

mediate product. Such complexity is required to guard the trading parties against changes in the market that might alter the precontract incentives to perform in the agreed manner. Market uncertainty increases transactions costs. In addition, reductions in the number of trading partners available to the firm are likely to increase the costs of market exchange because the firm will have few alternatives. The second set of factors is referred to as human factors (e.g., opportunism). Opportunism refers to situations where the use of deception can be expected to increase profits, and in such situations honesty in trading is not as likely.

Where the intermediate product is produced by the same firm that employs it, a considerable degree of flexibility to adapt to changing market conditions is realized. In addition, internal transfers are likely to reduce the opportunistic tendencies of the parties to the exchange. This results from the fact that "internal divisions [*see* INTERNAL ORGANIZATION OF THE FIRM] do not have preemptive claims on profit streams" (Williamson, 1974). The information flows between related stages of production are also likely to improve when these stages are combined within a single firm, for several reasons: there is a reduction in the incentive to behave opportunistically; common experiences tend to facilitate the overall flow of information within the firm; and the firm has much greater access to the relevant performance data of its internal divisions.

The key inquiry is whether vertical integration will result in increased profits. Several factors determine the answer to this query. These include:

- *Managerial capacity*. As the firm takes on more and more activities, managers lose track of things and the quality of managerial decisions suffers. The larger the firm, the longer the lines of communication between the manager and the production worker who must implement the decision. The capability of management limits the amount of information the manager can comprehend about the firm's operations. When the firm takes on additional functions, it can experience diseconomies (*see* DISECONOMIES OF SCALE) similar to those it experiences

when it expands output beyond the efficient scale of production.

- *Easily observable quality*. If an input is well defined and its QUALITY is easily determined at the time of purchase, it is more apt to be purchased in the market than produced internally, other things constant. For example, a flour mill will typically buy its wheat in the market while the manufacturer of a sensitive measuring instrument that requires a crucial gauge may produce it so that it can monitor quality. Another reason why producers sometimes integrate backward is so that they can make guarantees to their customers about the quality of the components or ingredients in a product (e.g., Frank Perdue raises his own chickens).

- *Number of suppliers*. A firm wants an uninterrupted source of component parts. When there are many interchangeable suppliers of a particular input, a firm is more likely to purchase that input in the market rather than produce it internally, other things constant. If the resource market is so unstable that the firm cannot rely on a consistent supply of the component, the firm may produce the item to insulate itself from the unstable market. In addition, the existence of a large number of potential suppliers increases the likelihood of COMPETITION between those firms driving down prices.

- *Monitoring*. The cost and ability to monitor production is another important factor. The more costly or difficult it is to monitor different production activities within a firm, the more efficient it may be to purchase components from external sources. This is related to some of the other factors mentioned above.

- *Contractual alternatives*. Various contractual arrangements exist that will result in a firm receiving some of the benefits exhibited by the vertically integrated firm. For example, a long term contract can resolve supply or demand reliability problems.

- *Legal limits on vertical integration*. Courts often scrutinize cases in which they feel a firm has vertically integrated to acquire MARKET POWER (Blair and Kaserman, 1983). The major objection of the courts involves vertical market foreclosure (BARRIERS TO ENTRY), which allegedly occurs when a supplier acquires one of its customers and this results in the supplier's rivals being prevented from competing for the acquired firm's business (Bork, 1969; Williamson, 1979). Both the SHERMAN ACT and the CLAYTON ACT have been used to challenge instances of vertical integration (*see* ANTITRUST POLICY (US)). The Department of Justice and courts are starting to focus their analysis of vertical integration cases on more than simply market foreclosure (Burns, 1993). The three competitive problems that can result from vertical integration identified by the Department of Justice are: increased entry barriers; possibility of COLLUSION; and the possible evasion of price regulation.

In summary, whether vertical integration will increase profitability or market power for a firm is not always obvious and depends on a slew of institutional, legal, and structural factors.

Bibliography

Blair, R. D. and Kaserman, D. L. (1983). *Law and Economics of Vertical Integration and Control*. Orlando, FL: Academic Press.

Bork, R. (1969). *Public Policy Toward Mergers*. Pacific Palisades: Goodyear.

Burns, J. W. (1993). Vertical restraints, efficiency, and the real world. *Fordham Law Review*, **62**, 597.

Coase, R. H. (1937). The nature of the firm. *Economica*, **4**, 386 405.

Schupack, M. B. (1977). The theory of vertical integration: A survey. Working paper no. 77-31.

Williamson, O. E. (1974). The economics of antitrust: Transactions cost considerations. *University of Pennsylvania Law Review*, **122**, 1439 96.

Williamson, O. E. (1979). Assessing vertical market restrictions: Antitrust ramifications of the transaction cost approach. *University of Pennsylvania Law Review*, **127**, 953.

X

X-efficiency

Kent A. Jones

X efficiency is a state of production in which, for a given level of input usage, there is no way of increasing the output. The output process is thus also said to be production efficient (Friedman, 1990: 436). In contrast, X inefficiency represents the failure to obtain maximum output from a given combination of inputs. Leibenstein (1966) developed the idea of X inefficiency as a way of describing the results of a weakened incentive structure for cost conscious, efficient management, especially in the absence of competitive market forces. For example, a state run MONOPOLY that does not incorporate the goal of PROFIT MAXIMIZATION tends to create an incentive structure among managers in the firm discouraging aggressive cost efficiency. In general, government bureaucracies driven by the goal of maximizing their operating budgets offer the most salient examples of X inefficiency (see Niskanen, 1971). Such managerial lassitude is based on the assumption that the market environment determines incentives for effort on the part of managers. Application of this concept to private monopolies or to firms enjoying government protection from COMPETITION is problematical, however (see Stigler, 1976). Critics of the concept of X efficiency claim that the existence of a monopoly or beneficial government intervention may not reduce managerial effort itself, but rather may merely redirect managerial efforts toward RENT SEEKING, for example. Nonetheless, to the extent that a weakening of the market discipline of competition reduces the incentive of managers to maximize the productive potential of inputs, technology, and research and development, the concept of X efficiency has some explanatory power.

Bibliography

Friedman, D. (1990). *Price Theory*, 2nd edn. Cincinnati, OH: South-Western Publishing.

Leibenstein, H. (1966). Allocative efficiency vs. "X-efficiency." *American Economic Review*, **56**, 392 415.

Niskanen, W. (1971). *Bureaucracy and Representative Government*. Chicago: Aldine-Atherton.

Stigler, G. J. (1976). The Xistence of X-efficiency. *American Economic Review*, **66**, 212 16.

Index