International Series in
Operations Research & Management Science

Giorgio Consigli
Silvana Stefani
Giovanni Zambruno *Editors*

Handbook of Recent Advances in Commodity and Financial Modeling

Quantitative Methods in Banking, Finance, Insurance, Energy and Commodity Markets

Springer

International Series in Operations Research & Management Science

Volume 257

More information about this series at http://www.springer.com/series/6161

Giorgio Consigli • Silvana Stefani
Giovanni Zambruno
Editors

Handbook of Recent Advances in Commodity and Financial Modeling

Quantitative Methods in Banking, Finance, Insurance, Energy and Commodity Markets

Springer

Editors
Giorgio Consigli
Department of Management, Economics and Quantitative Methods
University of Bergamo
Bergamo, Italy

Silvana Stefani
Department of Statistics and Quantitative Methods
University of Milan Bicocca
Milano, Italy

Giovanni Zambruno
Department of Statistics and Quantitative Methods
University of Milan Bicocca
Milano, Italy

ISSN 0884-8289 ISSN 2214-7934 (electronic)
International Series in Operations Research & Management Science
ISBN 978-3-319-61318-5 ISBN 978-3-319-61320-8 (eBook)
DOI 10.1007/978-3-319-61320-8

Library of Congress Control Number: 2017949871

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

A widespread liberalization process in commodity and energy markets has led over the last 15 years or so to a fruitful and rich methodological spreading of techniques and quantitative approaches previously proposed in financial markets into a wider global market area. At the same time, the increasing volatility of international prices and the introduction of regulatory frameworks on banking and insurance institutions enhanced the research on risk theory and risk management inducing new, practically relevant, theoretical developments. This handbook, at the time it was proposed to Springer, aimed at elaborating on such evidence to include contributions related to optimization, pricing and valuation problems, risk modeling, and decision-making problems arising in nowadays global financial and commodity markets from the perspective of operations research and management science.

The volume is structured in three parts, emphasizing common methodological approaches arising in the areas of interest:

1. Risk modeling
2. Pricing and valuation
3. Optimization techniques

Our original aspiration, as volume editors, was to collect within such structure a comprehensive set of recent state-of-the-art and original works addressing a variety of management and valuation problems arising in modern financial and commodity markets, such as:

- Risk measurement methodologies, including model risk assessment, currently applied to energy spot and future markets and new risk measures recently proposed to evaluate risk-reward trade-offs in global financial and commodity markets.
- Decision paradigms, in the framework of behavioral finance or factor-based or more classical stochastic optimization techniques, applied to portfolio selection problems including new asset classes such as alternative investments.

- Derivative portfolio hedging and pricing methods recently put forward in the professional community in the presence of increasing instability in financial as well as commodity markets.
- The adoption of multi-criteria and dynamic optimization approaches in financial and insurance markets in the presence of market stress and growing systemic risk.

Upon volume completion, we may say that most of the original research objectives have been reached and the 14 chapters included in this volume span a large and diversified variety of modeling and decision-making problems with a range of underlying methodological implications. We eventually decided to structure the content putting first the chapters primarily concerned with risk modeling and risk assessment issues, then those proposing (risk) pricing techniques, and finally those focusing on optimal risk control and decision-making paradigms.

Part I of this volume, on *risk modeling*, includes five chapters. The first chapter by Malliaris–Malliaris focuses on the market dynamics of gold and silver as commodities and analyzes in particular the directional predictability of their daily returns. The authors propose an interesting application of cluster analysis leading to the identification over a 15-year period of six important clusters, whose evaluation allows the definition of strategies within this market of precious metals. Three strategies in particular are evaluated which establish a relevant evidence for directional strategies in commodity markets, based on their lagged negative correlation: gold appears leading silver movements with a stable anti-correlated dynamics. The role of commodities in global financial portfolios has been advocated for their importance in enhancing real, inflation-adjusted returns and also due to their diversification gain relative to fixed-income and equity investments. Here the authors emphasize that indeed, also within commodity markets, investors and financial agents can profit from the commodity diversified market dynamics and their relationship with the business and economic cycle.

Sarwar et al. focus in Chap. 2 on credit-rated stocks and analyze how indeed a different approach to investment-grade rather than speculative-grade equities may generate significant momentum returns across business cycles with evidence of anti-cyclical patterns. During the period 1985–2011, the authors analyze in detail the US market and report that momentum returns from speculative-grade stocks amount on average to 1.27% per month and are more prevailing during contraction periods, in which they earn 1.61% per month. Furthermore investment-grade stocks are found to earn, on average, momentum returns of 0.85% per month and 1.14% per month during contractions. Momentum returns are in general associated with trading strategies based on canonical buy/sell signals associated with recent past winners vs. past losers, respectively. Interestingly, during the 2008 crisis, higher momentum returns are not explained by macroeconomic variables. The authors' overall conclusion is that positive momentum returns are due to high uncertainty associated with the increased credit risk of stocks and across business cycles. Such conclusion provides evidence of a persistent excess risk premium in speculative markets, with companies that in trouble periods either consolidate their business or go bankrupt.

In the third chapter, Sannajust–Chevalier analyze from a different perspective a developing equity market such as the emerging Asian private (rather than exchange-based or public) one, focusing on leveraged buyout (LBO) operations and their correlation with the target companies' performance over short and long term. The research spans a large set of candidate drivers (financial, governance, macroeconomic, cultural, microeconomic, and industry variables), and the authors base their analysis on the Capital IQ database. They focus in particular on the impact of macroeconomic factors on the performance of LBOs in Asia during the first decade of this century. The study, thus extending previous evidence on developed markets, shows that GDP growth, industry growth, and market return are important drivers that significantly contribute to create value in LBOs. It is worth recalling that, over the last 15 years, the private equity market attracted increasing interest due to the stable excess performance produced in the long term by this market and its increasing role as vehicle to attract equity investors at a time in which fixed-income returns were decreasing in developed as well as developing markets and financial instability and systemic risk were increasing.

D'Ecclesia–Kondi in the following Chap. 4 provide an interesting and in-depth methodological survey of the state of the art on correlation assessment methods across financial and commodity markets: as is well known, correlations between different asset returns represent a crucial element in asset allocation decisions as well as exotic derivative pricing. In commodity markets where prices are reported to be mostly nonstationary and returns are only mean stationary, a time-varying measure of correlation is needed, and indeed it is such assumption that in the first place leads to the emergence of correlation clustering phenomena during turbulent market phases. According to the prevailing literature, correlations among different markets are known to be higher during recessions than during expansion periods. When applied to portfolio management, with an investment universe including both financial and real assets, in order to shield investors from equity declines, portfolio managers historically used to invest in commodities deemed poorly correlated with stock markets. The authors clarify in their study, with an extensive data analysis, that during the last decade, also due to an increasing speculative role of many commodities, correlations between commodities and stock returns have dramatically changed and an accurate risk assessment may no longer be attained without introducing a correlation model assuming nonstationary data and structural breaks in market variables. The authors compare the historical rolling correlation and the dynamic conditional correlation methods and show how each estimator can provide useful information given a specific data structure and that information provided by the correlation measures can be used to identify structural breaks in the original variables. The analysis performed by D'Ecclesia and Kondi contributes, albeit indirectly, to underline the relevance of the adopted correlation model in the solution of a generic allocation problem.

In the fifth chapter, Gianfreda–Scandolo address directly the issue of measuring the cost generated by a wrong model. Indeed it has been shown that model risk has an important effect on any risk measurement procedures; therefore, its proper quantification is becoming crucial in several application domains. The authors analyze in particular the case of energy markets, where traders and market participants face several kinds of risks including market, liquidity, and, more importantly, operational risk. The authors propose the assessment of model risk in the German wholesale electricity market, looking at daily spot prices and comparing several models presented in the literature with their possible variations. Gianfreda and Scandolo propose a quantitative measure of model risk, namely, the relative measure of model risk, as proposed by Barrieu and Scandolo (2015). They quantify the model risk by studying day-ahead electricity prices in the European Energy Exchange (EEX). Germany, indeed, decided to exit from nuclear power by 2020 focusing on renewable energy sources and energy efficiency. This market is characterized by a high wind penetration which has increased the complexity of the electricity price dynamics given that wind (and solar) energy is highly variable and partially predictable. Model risk assessment is in this study applied to a specific energy market, but the research over possible quantitative methods to measure the impact of inaccurate or even wrong model assumptions on pricing, as well as risk management and decision models, is ongoing and attracting increasing interest, also through the so-called model sensitivity analysis as well as counterfactual analysis in commodity and financial markets. The topic is indeed becoming a specific task of many risk management units in global financial institutions and investment banks.

Part II on *pricing and valuation* collects contributions in which new and valuable techniques are introduced and described for pricing and evaluating financial products. This part includes four chapters in which the prevailing research focus is on pricing and calibration methods mainly in derivative markets with again as in Part I a variety of underlying assets, commodity or financial.

Noparumpa et al. provide in Chap. 6 a thorough analysis of the market of wine (mainly US) futures and the determinants of price formation and decision-making by wine producers taking into account spot vs. future price dynamics (their basis risk). The authors move from a detailed study of the determinants of wine prices and their dependence on seasonal and quality uncertainty to consider the drivers of price settlements in spot and future markets. This agricultural market represents a large and growing share of agri-markets primarily in developed but increasingly in selected developing markets. The study takes into account wines with different aging and production methods to infer the producer's decisions on (1) the sale price of her/his wine futures, (2) the quantity of wine futures to be sold in advance, and (3) the amount of wine to be kept for retail and distribution. The study makes two contributions to the optimization of pricing and quantity decisions by wine managers. A stochastic optimization model that integrates uncertain consumer valuations of wine both in the form of futures and in bottle and the uncertainty associated with bottle scores is also proposed with a detailed empirical analysis based on data collected from Bordeaux wineries engaging in wine futures.

In a rather different setting, Hitaj et al. discuss in Chap. 7 the important (methodological thus general) problem of describing log return dynamics in option pricing problems. It is well known that financial time series, increasingly in the recent past, exhibit heavy tails, asymmetric distribution, and persistence and clustering of volatility. The authors propose a class of discrete-time stochastic volatility models, starting from the affine GARCH model and assuming that the conditional distribution of log returns is a normal variance–mean mixture. They develop a discrete-time stochastic volatility model in a simple way, obtaining a recursive procedure for the computation of the log price characteristic function at option maturity. Finally, option prices are obtained via Fourier transforms. The authors are able to extrapolate information from the VIX data and find a linear relationship between the variance dynamics and the VIX^2. Moreover, this model is able to generate time-varying skewness and kurtosis that standard GARCH models cannot reproduce. Again, the issue of model risk assessment and the implications brought about by model selection are considered as in Chaps. 4 and 5 of this volume. The signaling power of the VIX is confirmed in the research. The authors also investigate the ability of the proposed modeling approach to reproduce the behavior of European option prices on SPX index. The dynamic normal inverse Gaussian-based model provides more flexibility in capturing market dynamics especially in turbulent periods.

Under more general assumptions, linking to the previous chapter, the important problem of finding a sound calibration method for pricing purposes is also discussed in Chap. 8 by Lindström–Åkerlindh. Indeed, while there is an abundance of good option valuation models, far less attention has been given in the literature to the key statistical problem of calibrating those models to market data and thus validate the proposed approaches. Local volatility models fit often perfectly with in-sample data, but the performance with out-of-sample data is less satisfactory. It is widely acknowledged that often practical calibration methods adopted in the financial industry reduce to some kind of least squares minimization of the difference between the fitted and observed data. Several studies have shown however that the weighted least squares (WLS) technique is practically infeasible when the model complexity grows, while nonlinear filters or penalized WLS work much better. A recent approach, proposed by one of the two authors, is based on using a nonlinear filter with time-varying model parameters, leading to more robust estimates and better out-of-sample forecasts. However, some tuning matrices were introduced that had to be tuned manually. The contribution in this volume extends the proposed methodology in two different directions: first by deriving a statistical framework for the tuning matrices and second by extending the dynamics of the original method from one to three different types of parameter dynamics. The proposed methodology, applied to European call options, is evaluated on several sets of simulated data as well as on S&P 500 index options from 2004 to 2008. The results are encouraging and capture well the structure of the underlying process. This may lead to improved and more effective hedging and risk management.

LIBOR-based derivatives (swaps, caps, swaptions) are the most liquid derivatives traded in global financial markets. Due to their importance and popularity, swaption market quotations are often used for calibration of interest rate models. However, the calibration procedure involves the pricing of a large number of swaptions (different option maturities, swap tenors, and strikes); then an efficient algorithm is required here. Since a closed-form formula of swaption prices does not exist for many popular interest rate models, then several approximate pricing methods have been developed in literature especially for affine interest rate models. In Chap. 9, extending previous results, Gambaro et al. establish a lower bound which is based on an approximation of the exercise region via an event set defined through a function of the model factors. The resulting formula consists in the valuation of the option on the approximate exercise region and requires a single Fourier transform performed through the appropriate parameter. The proposed approximation has several advantages. Indeed, by providing a lower bound, the direction of the error is known a priori; it is very general and involves the computation of only one Fourier inversion, independently of the number of cash flows of the underlying swap. Finally, it can be used as a control variate to improve the accuracy of the Monte Carlo simulation method.

Part III on *optimization* includes contributions in which maximization or minimization approaches take a prominent role in order to establish the best investment policies based on specific concave utility or convex risk functions, respectively. This part includes five chapters addressing different decision problems, from canonical one-period portfolio selection to multi-period institutional asset-liability management and hedging problems.

Hitaj–Zambruno discuss in Chap. 10 the effects of diversification constraints on the optimal portfolio choices by using the Herfindahl concentration index. In order to determine the optimal investment strategies, they use the third-order Taylor expansion of the exponential utility function to account for skewness. In the empirical analysis, these strategies are compared with others in the "smart beta" class and for various values of the risk aversion coefficient. The authors' contribution extends the domain of static portfolio selection methods, allowing an interesting comparison analysis.

In Chap. 11, Sbuelz investigates the joint effect of default risk and systemic risk on the dynamic asset allocation strategies in a no-arbitrage continuous time setting. This is accomplished by describing the dynamics of two representative assets as diffusion–jump processes, one of which is exposed to systemic risk only and the other also to default risk: the problem is formulated as a maximization problem of the expected power utility of terminal wealth. A numerical example shows the viability of the proposed model in the presence of systemic risk and interestingly highlights, under the given assumptions, the influence of an agent's time horizon.

In the following Chap. 12, Benazzoli–Di Persio focus on the implications of market liquidity in stock markets. They determine the optimal sequence of transactions required to sell a given amount of stock in an illiquid market, in which the trading rate affects prices. Such market impact is modeled by combining two effects: a permanent one, assumed linear in the trading rate, and a temporary one, represented through a negative exponential. The objective is to minimize the risk-adjusted expected costs of the strategy, where the control variable is represented by the transaction flow through time: a closed-form solution is obtained using the Lambert W function.

The issue of liquidity is also considered as a key strategy driver by Consigli et al. in Chap. 13, in which the elements of a real-world asset-liability management model of an occupational pension fund are considered. By adopting a multistage stochastic programming approach, the authors report how, from an initial underfunded status, a pension fund manager brings the fund to a fully funded status under different perspective scenarios over a 20-year planning horizon. The authors extend previous methodological approaches based on scenario trees to an interesting combination of decision stages distributed over time to annual liquidity assessments in which however investment rebalancing is not allowed. The presence of liquid as well as illiquid instruments in the investment universe has become a characterizing feature of global portfolios in the quest of excess returns at a time of unprecedented poor fixed-income returns. This chapter describes also in detail the adopted methodological and modeling steps leading to the completion of an advanced decision support tool for asset-liability management purposes.

In the final Chap. 14, Kallio et al. also adopt a stochastic programming approach, which in this case is applied to a currency hedging problem familiar to companies operating at an international scale. After an extensive review of the exchange rate dynamic models and the formulation of hedging techniques, the authors employ a multistage stochastic programming technique to determine the optimal hedging policy, the one providing at the end of the planning horizon the best risk-reward trade-off: working on actual data, they show not only that in general the model is effective in limiting downside risk but also that in specific periods the optimized policy can indeed improve profits from currency management by as much as 20%, particularly when leverage strategies are adopted.

This volume, in this reflecting the wide spectrum implied by its title, includes a variety of valuation and methodological problems emerging in different operational contexts, from developing private equity markets in Asia to liquid derivative markets either on commodities or on equity stocks as underlyings to again commodity futures in precious metals or global portfolios by pension fund managers. The volume also includes a set of dedicated contributions, primarily methodological, focusing on model risk, correlations, and stochastic volatilities, whose role in jeopardizing long-established results in mainstream finance has been remarked by many authors in recent times.

Upon completion of the editorial work, the editors would like to acknowledge the cooperation of the contributing authors and the continuing and productive assistance of Springer to achieve and complete the work.

Bergamo, Italy — Giorgio Consigli
Milano, Italy — Silvana Stefani
Milano, Italy — Giovanni Zambruno
March 2017

Contents

Part I Risk Modeling

1 Directional Returns for Gold and Silver: A Cluster Analysis Approach 3
A.G. Malliaris and Mary Malliaris

2 Impact of Credit Risk and Business Cycles on Momentum Returns . 17
Sirajum Munira Sarwar, Sharon Xiaowen Lin, and Yaz Gülnur Muradoğlu

3 Drivers of LBO Operating Performance: An Empirical Investigation in Asia 41
Aurélie Sannajust and Alain Chevalier

4 Time Varying Correlation: A Key Indicator in Finance 69
Rita L. D'Ecclesia and Denis Kondi

5 Measuring *Model Risk* in the *European Energy Exchange* 89
Angelica Gianfreda and Giacomo Scandolo

Part II Pricing and Valuation

6 Wine Futures: Pricing and Allocation as Levers Against Quality Uncertainty 113
Tim Noparumpa, Burak Kazaz, and Scott Webster

7 VIX Computation Based on Affine Stochastic Volatility Models in Discrete Time 141
A. Hitaj, L. Mercuri, and E. Rroji

8 Optimal Adaptive Sequential Calibration of Option Models 165
Erik Lindström and Carl Åkerlindh

9 Accurate Pricing of Swaptions via Lower Bound 183
Anna Maria Gambaro, Ruggero Caldana, and Gianluca Fusai

Part III Optimization Techniques

10 Portfolio Optimization Using Modified Herfindahl Constraint 211
Asmerilda Hitaj and Giovanni Zambruno

11 Dynamic Asset Allocation with Default and Systemic Risks 241
Alessandro Sbuelz

12 Optimal Execution Strategy in Liquidity Framework Under Exponential Temporary Market Impact 251
Chiara Benazzoli and Luca Di Persio

13 Optimal Multistage Defined-Benefit Pension Fund Management 267
Giorgio Consigli, Vittorio Moriggia, Elena Benincasa, Giacomo Landoni, Filomena Petronio, Sebastiano Vitali, Massimo di Tria, Mario Skoric, and Angelo Uristani

14 Currency Hedging for a Multi-national Firm 297
Markku Kallio, Matti Koivu, and Rudan Wang

Part I
Risk Modeling

Chapter 1
Directional Returns for Gold and Silver: A Cluster Analysis Approach

A. G. Malliaris and Mary Malliaris

Abstract This paper considers the directional predictability of daily returns for both gold and silver. These two metals have had a long history behaving sometimes as complements and other times as substitutes. We use daily data from June of 2008 through February of 2015. The last 2 years were removed as a set for validation of the model and the remainder, almost 5 years, was used as training. A cluster analysis yields six important clusters. An evaluation of these clusters leads to the formation of three strategies for directional predictions – up or down—for both gold and silver returns. The results of this analysis suggest that each strategy has its own advantages: the first strategy suggests that gold returns can be predicted better than those of silver; the second strategy shows that predicting up for gold also means predicting down for silver and the final strategy confirms that predicting up for silver also validates predicting down for gold.

Keywords Gold • Silver • Directional Forecasting • Cluster Analysis • Neural Networks

JEL Classification C5, C18, G1

1.1 Introduction and Literature Review

Milton Friedman (1990) offers a detailed historical analysis of bimetallism. He argues that monetary systems throughout the recorded history were based on precious metals and in particular silver and gold. Silver was used much more than

A.G. Malliaris (✉)
Department of Economics and Department of Finance, Quinlan School of Business, Loyola University Chicago, Chicago, IL, USA
e-mail: tmallia@luc.edu

M. Malliaris
Department of Information Systems and Operations Management, Quinlan School of Business, Loyola University Chicago, Chicago, IL, USA
e-mail: mmallia@luc.edu

G. Consigli et al. (eds.), *Handbook of Recent Advances in Commodity and Financial Modeling*, International Series in Operations Research & Management Science 257,
DOI 10.1007/978-3-319-61320-8_1

gold in Europe, India and Asia since it is more abundant than gold. Gold was used for high-valued transactions. Authorities always had problems between the legal rate of exchange of one metal for the other because market conditions were never fixed in exactly the same ratio as proposed by the authorities. After the 1870s, U.S. and most European countries shifted to using only gold, leaving India and China as the only two large countries still preferring silver. After World War I, the link between gold and national currency diminished and in August 15, 1971, the connection between gold and the dollar in international transactions was abolished.

Aggarwal, Lucey and O'Connor (2015), give a comprehensive analysis of the gold and silver markets as precious commodities divorced from their long history as money. Both continue to be important commodities that play the traditional roles of hedging, arbitrage, speculation, market efficiency and portfolio investment. An earlier classic paper by Escribano and Granger (1998) analyzes the long-run relationship between gold and silver. This study has been recently extended by Baur and Tran (2014) who study the role of bubbles and financial crises in gold and silver during the period 1970-2011. These authors find that a co-integration relationship exists between gold and silver with gold prices driving the relationship. They also find that these results are influenced by both bubble-like episodes and financial crises. Ciner (2001) shows that the stable relationship between gold and silver has disappeared in the 1990s and Batten, Ciner and Lucey (2015) establish that precious metals markets are weakly integrated and that this degree of integration is time varying.

In this paper we consider a speculator who wishes to take a daily position in gold and or silver. This position can be to buy or sell in the cash market or in a futures contract. Unlike earlier studies that analyze long-run relationships between gold and silver our emphasis is a 1 day investment horizon. Is there an appropriate methodology that such a speculator can employ? We propose to use a cluster analysis during a period of about 5 years, identify and analyze certain clusters and use the results of such an analysis over a long period of 2 years to forecast directional returns. The benchmark to contrast our results will be the random walk paradigm with a 50/50 chance for up and down. Thus we contribute to the gold and silver literature by selecting a topic that has received little attention, namely short-term speculation and secondly by employing the novel methodology of cluster analysis.

1.2 Data Collection and Preparation

The data set for this study goes from June of 2008 through February of 2015 and consists of daily values for Gold, Silver, and a Gold Volatility Index. This period covers the beginning of the Global Financial Crisis, few months prior to the Lehman Brothers bankruptcy in mid-September 2008 to the a couple of months after the termination of the third round of Quantitative Easing in late 2014. This is a highly volatile period that is challenging to model the price behavior of both gold and silver.

The values for Gold were downloaded from the St. Louis Fed FRED database using the series GOLDPMGBD228NLBM, which is the Gold fixing price at 3:00 P.M. (London time) in the London Bullion Market, based in U.S. Dollars. The Silver data were sourced from www.quandl.com as the price per troy ounce set at the London Fixing by the London Bullion Market Association. The Gold Volatility Index is the CBOE Gold ETF Volatility Index©, GVZCLS, from the Chicago Board Options Exchange, also downloaded through FRED. All data points represent daily prices. After downloading all data sets, the calculated fields are formed, and then the sets are matched on date.

The last 2 years of the data set is removed to be used as a set for validation of the model. The first part of the set, corresponding to approximately 5 years of data, is used for training purposes. The training set, from 6/9/2008 through 2/22/2013, has 1154 rows of data, and the validation set, from 2/25/2013 through 2/25/2015, has 493 rows of data. Gold, Silver, and the Gold Volatility index are scaled to be between 0 and 1. For each series, we create the following calculated fields: percent change from yesterday to today, the direction the series moved from yesterday to today, and the number of Up movements in the last 5 days. In addition, we calculate the difference in the Gold and Silver scaled values. This gives us a total of 13 input variables. The target variables are the directions that Gold or Silver will move from today to tomorrow. The variable names, their roles, and a brief description of each are shown in Table 1.1.

The scaled prices of the three base variables for both the training and the validation sets are shown in Figs. 1.1 and 1.2. We see that the highest values for all three series occurred within the training set.

Table 1.1 Variables used in the models

Variables	Role	Description
GDirTp1	Target	The direction Gold will move tomorrow
SDirTp1	Target	The direction Silver will move tomorrow
Gscaled	Input	Gold value scaled between 0 and 1
GVolscaled	Input	Gold Volatility value scaled between 0 and 1
Sscaled	Input	Silver value scaled between 0 and 1
GPerChg	Input	Gold percent change yesterday to today
GVolPerChg	Input	Gold Volatility percent change yesterday to today
SPerChg	Input	Silver percent change yesterday to today
Gdir	Input	The direction Gold moved from yesterday to today
GVoldir	Input	The direction Gold Volatility moved from yesterday to today
Sdir	Input	The direction Silver moved from yesterday to today
GDaysUp	Input	Number of Up moves for Gold in last 5 days
GVolDaysUp	Input	Number of Up moves for Gold Volatility in last 5 days
SDaysUp	Input	Number of Up moves for Silver in last 5 days
GscMinSsc	Input	Gold scaled value minus Silver scaled value today

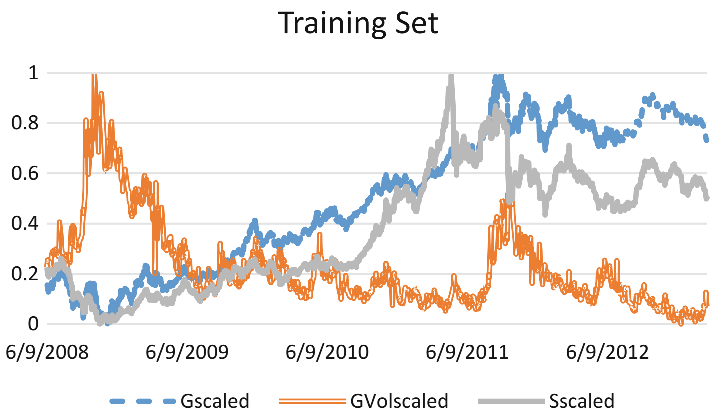


Fig. 1.1 Training set

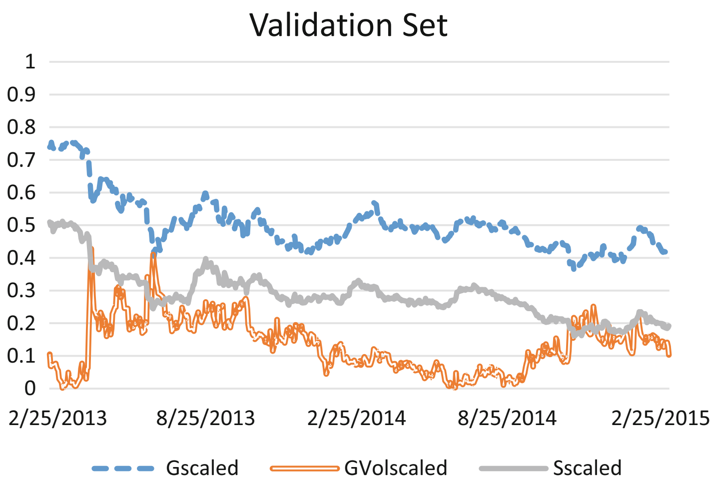


Fig. 1.2 Validation set

Of the derived variables created for this study, one is the number of day each of the base variables had been up in the last 5 days. That is, for about a week, how strong is the upward trend? We see these results in Fig. 1.3. Even though the training set has more values, we do note a similar spread of values across both sets. Proportionally, we observe that the training set has a higher number of equivalent values of 2 and 3 days up while the validation set has a higher number of 2 days up values than that of 3 days up.

In Fig. 1.4, we see the number overall of Down, Even and Up days in each of the sets. Again, we notice similar proportions across both sets. However, in Gold, there are more Up days in the training set. In Silver there are fewer Up days in the validation set, proportionally.

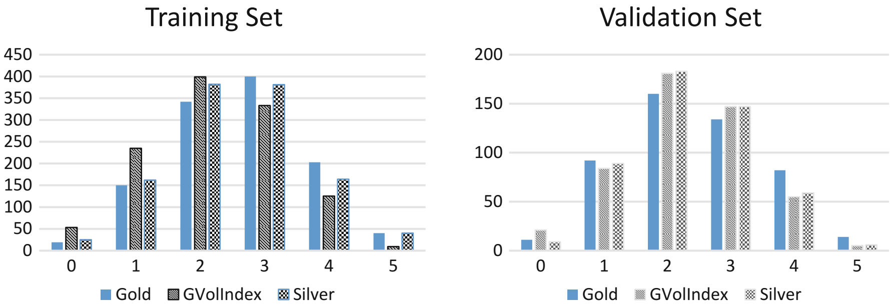


Fig. 1.3 Training and validation sets, number of days up in five

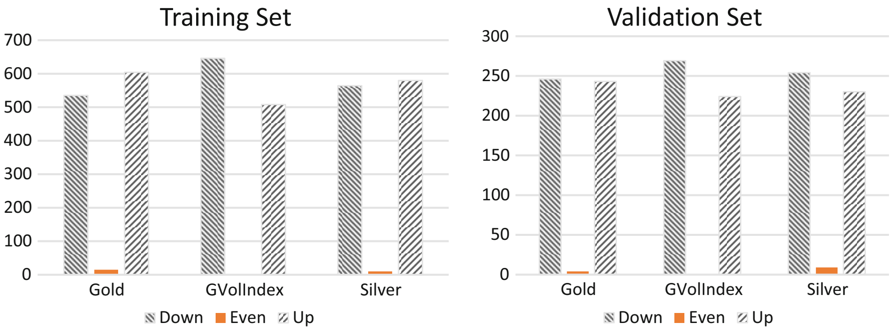


Fig. 1.4 Number of days in each direction

1.3 Methodology: Two-Step Cluster Analysis

After the construction of the training and validation sets, the next step is to generate clusters on the training set data. IBM's SPSS Modeler data mining software is used for this step. The objective in cluster analysis is to generate grouping of data where the rows within one group are more similar to each other than they are to rows in another group. Since you rarely know, in large data sets, the optimal number of groups to form, the cluster methodology used here, Two-Step, tests each possible configuration between two and fifteen groups. A silhouette measure of cohesion and separation is calculated for each group. This measure is not meaningful in isolation, but only in comparison to the measures generated by each of the possible configurations. The configuration with the best measure of cohesion and separation determines the final number of groups.

The Two-Step cluster analysis methodology can use both numeric and categorical inputs. It does not use a target variable, but forms the clusters only on the basis of the input variables. It processes the data in two steps. In the first step, it forms a large number of small sub-clusters that occur naturally within the data set. In the second step, it joins similar small sub-clusters together to make larger clusters.

Table 1.2 Cluster size and description

	Cluster-1	Cluster-5	Cluster-2	Cluster-6	Cluster-3	Cluster-4
Cluster size in rows	199	203	232	186	151	183
GDaysUp	3.2	2.21	3.13	2.02	2.23	2.85
Gdir	U 100%	D 100%	U 98.7%	D 99.5%	D 96.7%	U 96.2%
GPerChg	0.01	−0.01	0.01	−0.01	−0.01	0.01
GScMinSSc	0.11	0.11	0.13	0.14	0.12	0.13
GVolDaysUp	2.73	1.88	1.83	2.78	2.11	2.15
GVolDir	U 100%	D 100%	D 100%	U 100%	D 60.9%	D 65%
GVolPerChg	0.05	−0.04	−0.03	0.05	−0.01	−0.01
SDaysUp	3.04	2.13	3.00	2.01	2.99	2.01
Sdir	U 99.5%	D 99.5%	U 100%	D 99.5%	U 99.3%	D 96.2%
SPerChg	0.02	−0.02	0.02	−0.02	0.01	−0.01

This methodology uses a log-likelihood distance measure, with a probability-based distance. The distance between any two clusters is related to the decrease in log-likelihood as they are combined into one cluster. For details and formulas relating to the cluster formation, see the IBM SPSS Modeler 16 Algorithms Guide (2013).

Using the Two-Step methodology, it is determined that the optimal number of clusters in this training data set is six. These six clusters ranged in size from 151 to 232 rows. A cluster analysis creates a new column giving, for each row in the data set, an assignment of the cluster to which the row belongs. Since the clusters are created using only input variables, we can also run the validation set, or any new future set, through the trained cluster model to get cluster assignments for these new rows. The size and description of each of the clusters are given in Table 1.2. For each of the input that is considered important by the methodology, the average (for numeric variables) or the mode (for category variables) is used as a descriptive picture of the cluster. These clusters have three basic types based on the movements of Gold, Silver, and the Gold Volatility Index:

In the first type, all three base variables moved Up. This occurred in clusters 1 and 5.
The second type, seen in clusters 2 and 6, has Gold and Silver moving in the same direction, but the Gold Volatility Index going the opposite way.
Last, we see two clusters where Gold and Silver moved in opposite directions with Gold Volatility always Down. This occurred in clusters 3 and 4.

Using the information from the cluster analysis, we build separate decision trees on the training set for each of these cluster groupings (1 & 5, 2 & 6, and 3 & 4). We then evaluate the models by looking at their performance on the 2-year validation set. The cluster number is added as an input into the models when training.

The decision tree approach selected for this analysis is the C5.0 algorithm. The algorithm runs with Modeler, but is licensed from RuleQuest. Details can be found at the RuleQuest website at http://www.rulequest.com/. A decision tree algorithm uses both input variables and a target variable. In the C5.0 algorithm, the target

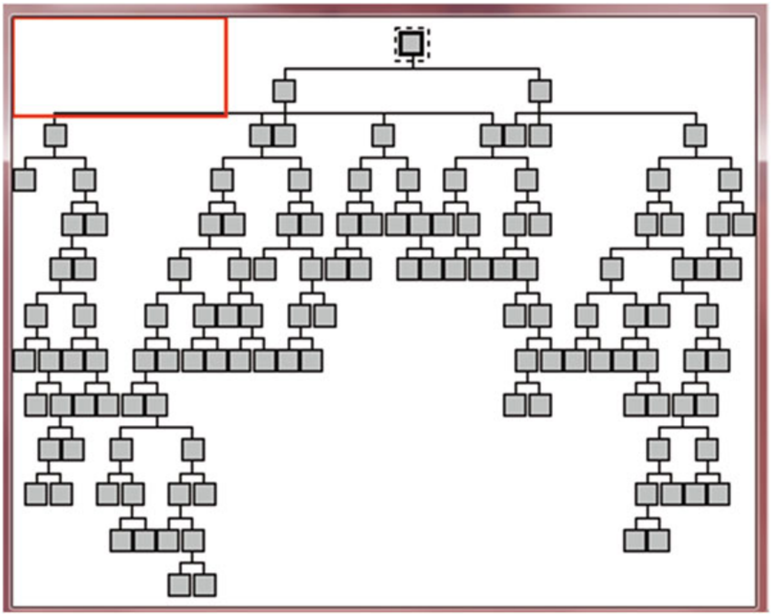


Fig. 1.5 Example decision tree generated by the C5.0 methodology

variable is a category-type variable, in this case the direction that Gold will move tomorrow. (For the second part of our study, this target is changed to Silver's direction tomorrow.) The input consists of the input variables listed in Table 1.1 plus the cluster assignment generated by the Two-Step algorithm.

Using the target variable, the C5.0 methodology builds a set of decision rules that determine the way each row is assigned a final value of the target variable. The decision tree begins with all the data in one set. It tests each input variable to see which single variable splits the training set into the most pure subsets of the target variable. Using the single best variable, a split of the training set is formed that gives two or more subsets, with each subset having a dominant single value of the target. The initial process is repeated on these new subsets. That is, each input variable is tested to see which one would optimize the purity of the target variable in subsets generated by a split based on that variable. This process continues on each subset until one of two things happen: Either the subset has a single value of the target, or there is no input variable split that can improve the purity. An optimal solution occurs when each subset is single-valued on the training set.

The C5.0 procedure generates both a tree-shaped output and a set of rules corresponding to each path in the tree. Figure 1.5 illustrates one decision tree. The node at the top, the root node, contains all the rows of data. As shown in the figure, this data is split into nodes one level below that contain a more pure division of the data on the values of the target variable. Each of the nodes at this level is then split further. Each level shows nodes that split further into another level, or that end.

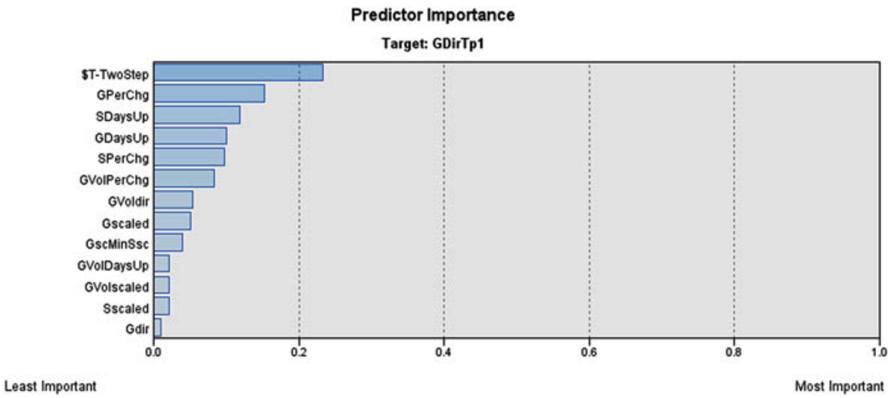


Fig. 1.6 Relative importance chart of each input in a decision tree, generated by Modeler

Once a node stops splitting, the C5.0 methodology generates a rule that corresponds to the path from the root node to that specific end node. An example rule might be "If SPerChg <= 0.027 and $T-TwoStep in Cluster-1 and GVolDaysUp > 2 and GVolscaled <= 0.772 and GVolPerChg <= 0.145 and GVolDaysUp <= 3 and SDaysUp > 1 and GDaysUp <= 2 THEN GDirTp1 = D". Each rule ends with a predicted value for the target variable. There will be as many rules as there are ending nodes. Also, for any set of values of current or future rows of data, there will be some path through the tree that corresponds to that data. Once the rules are generated, they can be applied to future input data to generate predicted values for the target variable.

Each trained decision tree model also yields a set of variable relative importance values that give us an indication of how each variable was finally valued by the model. These relative importance values sum to 1, and larger values indicate higher impact in determining the model forecast. Figure 1.6 shows an example of a graph of these relative importance values.

When a variable is not used at all by the model, its impact has a value of zero. In considering these values, there is no specific level of value by which we measure importance, or lack of importance, to a model. Rather, we use the values simply to compare within a model how that particular model used the input to form its predictions.

1.4 Gold with Clusters

1.4.1 Training Set Variable Importance

Three C5.0 decision tree models are built to forecast the direction Gold will move tomorrow, one tree for each of the cluster groupings. Table 1.3 gives the relative importance of the input variables for each of the three decision tree models built to forecast Gold. We see that each of the models valued the input variables in a different way. The numbers in this table reflect the relative importance of each input variable in the specific model. The three variables with the most impact for each model are shown as shaded fields.

Focusing on these three top variables for each model, we see that each model has at least one variable from the Gold set and one from the Gold Volatility Index set of derived variables. Two of the models show a high value on one of the Silver variables while the model built on clusters 2 & 6 does not value any of the Silver variables highly. In the model for clusters 1 & 5, the single variable with the most impact is the scaled value of the Gold Volatility Index. In the model for clusters 2 & 6, the most important single variable is the percent change in Gold, and in the model for clusters 3 & 4, it is the scaled value of Gold. Thus, we see that these models are distinctly different in their approach to making a decision about tomorrow's movement in Gold. No single model would do as well as one built on separate clusters.

One way to judge the cumulative effectiveness of the three most important input variables in each column is to add the relative importance of the three most important variables in the three categories of clusters as presented in the last row of Table 1.3. This last row implies that the cumulative impact of the three most important variables differs in the three columns with the last column being the most significant.

In building a model for the future, the basis for judging is how well it performs on the validation set of data that the model was not trained on. The next section demonstrates the performance of our model by applying the findings from the training data set to the validation data set.

Table 1.3 Relative importance of each input variable to the decision tree model

Variables	Importance for Clusters 1 & 5	Importance for Clusters 2 & 6	Importance for Clusters 3 & 4
GDaysUp	0.1321	0.0795	0.0022
Gdir	0	0.111	0.0485
GPerChg	0.0464	0.3841	0.0287
Gscaled	0	0.0194	0.5062
GVolDaysUp	0.0442	0.1645	0.0418
GVoldir	0	0	0.0695
GVolPerChg	0.108	0.0531	0.0461
GVolscaled	0.2811	0	0.0509
SDaysUp	0.089	0.0469	0.0008
Sdir	0.0391	0	0.008
SPerChg	0.0565	0.088	0.1607
Sscaled	0.1691	0.0212	0.0113
GscMinSsc	0.0344	0.0322	0.0254
Cluster Number	0	0	0
Cumulative Impact of 3 top variables	.5823	.6596	.7364

1.4.2 Validation Set Results for the Gold Models

The validation set consists of the 2 years following the training set. Each row is run through the trained cluster model to generate a cluster number. Following the cluster assignment, rows are fed through the appropriate trained decision tree model to generate a forecast. Generated forecasts are then compared to the actual directional movement of Gold on the following day. Results are shown in Table 1.4.

We see that there are very few days with no change (even) and none of these are identified by any model since the models were trained to recognize only Up and Down days. The model for clusters 1 & 5 predicts Down 35 times of which 20 were correct (57.1%). It predicts Up 113 times with 66 of these correct (58.4%). The model for clusters 2 & 6 is correct in predicting Down 62 times out of 118 (52.5%) and correct in the Up forecasts 44 of 72 times (61.1%). In the last model, for clusters 3 and 4, there are 39 correct Down predictions out of 63 (61.9%) and 44 of 92 correct Up predictions (47.8%).

Combining the correct numbers over all directions in Table 1.5, we see that the percent of times that the forecast matches the actual direction is about 54% in the worst model and 58% in the best model. Forecasting over a 2-year time period without retraining is a difficult task for any model. That these models are able to remain as correct as they are for this extended time is an indication that they discover some stable rules in their training sets.

Table 1.4 Validation set results, correct forecasts shown in bold

	Actual direction of gold tomorrow				Values as proportion of total forecasts		
Forecasted direction	Down	Even	Up	Total forecasts	Down	Even	Up
Clusters 1 & 5							
Down	**20**		15	35	**0.571**		0.429
Up	47		**66**	113	0.416		**0.584**
Clusters 2 & 6							
Down	**62**	3	53	118	**0.525**	0.025	0.449
Up	28	0	**44**	72	0.389		**0.611**
Clusters 3 & 4							
Down	**39**	0	24	63	**0.619**		0.381
Up	47	1	**44**	92	0.511	0.011	**0.478**

Table 1.5 Overall proportion of correct forecasts

Model for:	Proportion of correct forecasts
Clusters 1 & 5	0.581
Clusters 2 & 6	0.567
Clusters 3 & 4	0.539

1.5 Silver with Clusters

1.5.1 Training Set Variable Importance

As with the Gold forecasts, three separate decision tree models are trained, one for each of the cluster groupings for Silver. For these decision tree models, trained to forecast the direction Silver will move tomorrow, the input consists of the variables listed in Table 1.1 plus the cluster assignment variable. All decision trees use the C5.0 methodology and are built using the IBM SPSS Modeler 16 software.

The variable importance results are shown in Table 1.6 and the three most important variable values of each model are shaded. The cluster number turns out to be an important variable in the cluster 3 & 4 model. This is the only place, in either the Gold or the Silver models, that the specific cluster assignment plays an important role. We see that Silver derived variable ranks highly in the models for clusters 1 & 5 (Sdir) and for clusters 2 & 6 (SDaysUp). In these two models, of the two other most important variables, one is based on Gold and one on the Gold Volatility Index. So some derived variable from each of the base variables plays an important part in the decision tree rules. The most important variables for the model built on clusters 3 & 4 are two variables based on Gold, and the cluster number itself. Both Silver and the Gold Volatility Index have more minor roles in this model.

Table 1.6 Relative importance of input variables for Silver forecasts

Variables	Importance for Clusters 1 & 5	Importance for Clusters 2 & 6	Importance for Clusters 3 & 4
Cluster Number	0	0	0.2272
GDaysUp	0.1226	0.0523	0.0637
Gdir	0	0.5006	0.3153
GPerChg	0.1286	0.0111	0.1387
Gscaled	0.0254	0.0468	0.0838
GVolDaysUp	0.182	0.0324	0.0158
GVoldir	0	0	0.0695
GVolPerChg	0.0757	0.0858	0.0014
GVolscaled	0.0208	0.0262	0.0061
SDaysUp	0.0864	0.1186	0.0218
Sdir	0.1866	0	0
SPerChg	0.0586	0.0811	0
Sscaled	0.0515	0.0404	0.0176
GscMinSsc	0.0617	0.0047	0.039

Table 1.7 Validation set results, correct forecasts shown in bold

	Actual direction of silver tomorrow				Values as proportion of total forecasts		
Forecasted direction	Down	Even	Up	Total forecasts	Down	Even	Up
Clusters 1 & 5							
Down	**32**	0	34	66	**0.485**	0	0.515
Up	40	2	**40**	82	0.488	0.024	**0.488**
Clusters 2 & 6							
Down	**70**	2	36	108	**0.648**	0.019	0.333
Up	41	3	**38**	82	0.500	0.037	**0.463**
Clusters 3 & 4							
Down	**33**	0	29	62	**0.532**	0	0.468
Up	38	2	**53**	93	0.409	0.022	**0.570**

Table 1.8 Overall proportion of correct forecasts

Model for:	Proportion correct
Clusters 1 & 5	0.486
Clusters 2 & 6	0.568
Clusters 3 & 4	0.555

1.5.2 Validation Set Results for the Silver Models

After running the validation set through the appropriate trained model, the number of correct forecasts for each model are shown in Table 1.7.

As with the models for Gold, we find that no model makes any forecast for the Even direction. This is due to the fact that there are so few Even days that the models cannot find stable rules to apply. Here we see that the models built on clusters 1 & 5 do not perform better that 50% for either the Down or the Up direction. The model for clusters 2 & 6 perform well in forecasting Down, but not well on Up. The model for clusters 3 & 4 results in a performance above 50% on both Down and Up forecasts, with a better performance on Up.

Combining the correct numbers over all directions in Table 1.8, we see the percent of times that the forecast matches the actual direction. This is approximately 48% in the worst model and 57% in the best model. However, because of the very different results in forecasting Down and Up, it is essential to pay attention to direction in the case of Silver. When a model fails to remain robust on the validation set, it is often an indication that the underlying patterns in the relationships have altered after the training data is completed.

Comparing the overall correctness for Gold forecasts and for Silver forecasts, we see, in Table 1.9, that on clusters 1 & 5, the model for Gold outperformed the model for Silver. For the model on clusters 2 & 6, both performed similarly. And for the models built on clusters 3 & 4, Silver slightly outperforms Gold.

Table 1.9 Comparison of Gold and Silver forecasts

Model for:	Gold correct	Silver correct	Gold minus silver
Clusters 1 & 5	0.581	0.486	0.095
Clusters 2 & 6	0.567	0.568	−0.001
Clusters 3 & 4	0.539	0.555	−0.016

Table 1.10 Comparison of directions, Gold and Silver models

Forecasted direction	Down	Up
Clusters 1 & 5		
Down	Gold	
Up		Gold
Clusters 2 & 6		
Down	Silver	
Up		Gold
Clusters 3 & 4		
Down	Gold	
Up		Silver

Drilling down to specific directions, Table 1.10 illustrates the performance of the model on each set of clusters. For the models built on Clusters 1 & 5, the Gold model performs better predicting both Down and Up days than do the Silver models. For the models built on clusters 2 & 6, the Silver models predict Down better and the Gold model predicts Up better. On the clusters 3 & 4 models, a reversal is observed with the Gold model performing better on Down forecasts and the Silver model performing better on Up days.

1.6 Summary and Conclusions

Unlike earlier studies that approach gold and silver relationships over very long periods using time series techniques, this paper focuses on the gold and silver daily directional returns using the Two-Step cluster methodology as a beginning point, with C5.0 decision trees built on similar groups of clusters.

We derive data based on daily values of Gold, Silver and the Gold Volatility Index to ultimately forecast the direction that gold and that silver will move tomorrow. These forecasts are done with decision trees using a training set of almost 5 years and a validation set of 2 years. The cluster analysis, based only on the training set, finds six clusters in the base data and assigns a cluster number to each row. These six clusters are combined into groups of two based on the behavior of the base variables. Each decision tree is built on a specific pair of clusters that show some similarity of behavior. Separate groups of decision trees are generated for gold and for silver forecasts.

We find that, by combining six clusters to form three distinctive strategies, we can outperform the random walk 50/50 up or down prediction. The clusters show that gold and silver follow each other closely but often deviations occur where one price goes up while the other goes down with interchanging leadership in various patterns.

References

R. Aggarwal, B. Lucey, F. O'Connor, World metal markets, in *The World Scientific Handbook in Futures Markets*, ed. by A. G. Malliaris, W. Ziemba (Eds), (World Scientific Publishing, Singapore, 2015)

J. Batten, C. Ciner, B. Lucey, Which precious metals spill over on which, when and why?—Some evidence. Appl. Econ. Lett. **22**, 466–473 (2015)

D. Baur, D. Tran, The long-run relationship of gold and silver and the influence of bubbles and financial crises. Empir. Econ. **47**, 1525–1541 (2014)

C. Ciner, On the long run relationship between gold and silver prices: a note. Glob. Financ. J. **12**, 299–303 (2001)

A. Escribano, C.W. Granger, Investigating the relationship between gold and silver prices. J. Forecast. **17**, 81–107 (1998)

M. Friedman, Bimetallism revisited. J. Econ. Perspect. **4**(4), 85–104 (1990)

IBM SPSS Modeler 16 Algorithms Guide. (IBM Corporation, 2013), pp. 413–417

Chapter 2
Impact of Credit Risk and Business Cycles on Momentum Returns

Sirajum Munira Sarwar, Sharon Xiaowen Lin, and Yaz Gülnur Muradoğlu

Abstract In this paper, we show that significant momentum returns generate from credit-rated stocks across business cycles. The generation of momentum earned from speculative-grade stocks is on average 1.27% per month and are more prevalent during contraction periods in which they earn 1.61% per month. We also find that investment-grade stocks earn on average momentum returns of 0.85% per month and 1.14% per month during contractions. Higher momentum returns are unexplained by macroeconomic variables during contractions such as the 2008 recession. Our findings conclude that momentum return is due to high uncertainty associated with the increased credit risk of stocks and across business cycles.

Keywords Credit-rated stocks • Business cycles • Momentum • Uncertainty

JEL Classification G11, G12, G19

We thank Richard Fairchild, Nigel Harvey, David Hirsleifer, Amy Kam, Meziane Lasfer, Mario Levis, Lucio Sarno, Richard Taffler, and conference and seminar participants at the FMA in Europe in Prague, EWGFM in London, Sir John Cass Business School, Ted Rogers School of Management, Westminster Business School, Newcastle Business School, and Keele University. All errors are our own.

S.M. Sarwar
Bentley University, Office: Morison 178, 175 Forest Street, Waltham, MA, 02452, USA
e-mail: ssarwar@bentley.edu

S.X. Lin
NIHR CLAHRC Wessex Data Science Hub, Faculty of Health Sciences, University of Southampton, Southampton General Hospital (Room AA72, MP11), Southampton, SO16 6YD, UK
e-mail: X.Lin@soton.ac.uk

Y.G. Muradoğlu (✉)
School of Business and Management, Queen Mary, University of London, Francis Bancroft Building, Mile End Road, London, E1 4NS, UK
e-mail: y.g.muradoglu@qmul.ac.uk

G. Consigli et al. (eds.), *Handbook of Recent Advances in Commodity and Financial Modeling*, International Series in Operations Research & Management Science 257,
DOI 10.1007/978-3-319-61320-8_2

2.1 Introduction

It is well-established that momentum returns, that result from the trading strategy of buying recent past winners and selling recent past losers, earn statistically and economically significant profits of 1% per month or 12% per annum (Jegadeesh and Titman 1993, 2001). Subsequent studies find both risk-based explanations (e.g., see Eisdorfer 2008; Avramov et al. 2007; Cooper et al. 2004; Chordia and Shivakumar 2002; Harvey and Siddique 2000; Jegadeesh and Titman 1993, 2001) and behavioural models (e.g., see Chui et al. 2010; Korajczyk and Sadka 2004) that explains momentum phenomenon. Despite this progress, the persistence of momentum returns remains robust.

Recent studies on risk-based models report that momentum is observed more in stocks with high information uncertainty, default risk, in periods of high market volatility and stocks that are credit rated (e.g., see Avramov et al. 2007; Bhar and Malliaris 2011; Jiang et al. 2005; Wang and Xu 2009; Zhang 2006; Lee 2012). Avramov et al. (2007) show momentum returns are high among low-grade firms and are nonexistent among high-grade firms. The findings of the study by Avramov et al. (2007) imply that momentum returns should be higher during recessionary periods when credit risk is high. However, their time series analysis indicates otherwise. This is puzzling. They also advise that "future work should address" this issue (Avramov et al. 2007, p. 2520). In this paper, we do that. We show that momentum returns are earned by speculative-grade stocks and investment-grade stocks during recessions, but the returns are more pronounced in speculative-grade stocks. Speculative-grade stocks carry high uncertainty in terms of company prospects. During recessions, credit risk is a major concern and imposes additional uncertainty. Momentum returns compensate for both the credit risk of a company and the state of the business cycle.

We contribute in the literature by focusing on the behaviour of different types of credit rated stocks across business cycle. We differ from the previous study of Avramov et al. (2007) that we have divided stocks into two broad category of investment grade and speculative grade stocks, the two groups of stocks that the investors are interested to invest. The purpose is to study the generation of momentum returns of these two groups of stocks, e.g., investment-grade and speculative-grade. Therefore, we divide credit-rated stocks into two categories, investment-grade and speculative-grade, to understand the behaviour of momentum returns in these categories. Investment-grade stocks have low credit risk and thus low uncertainty. Speculative-grade stocks have higher credit risks and thus higher uncertainty. Their default rates are as high as 6.53%. We find that momentum returns on average are 0.85% per month for investment-grade stocks while they are 1.27% per month for speculative-grade stocks. Momentum studies document that momentum profits have started to disappear, the process that began in the early 1990s, was only delayed by the tech-boom, and then faded away afterwards (e.g., see Hwang and Rubesam 2008; Wang and Xu 2009; Bhattacharya et al. 2011). We observe momentum returns in the US market during the tech-boom of the 2000s and the subprime financial crisis of 2008. We report that in the early 2000s investment-grade stocks earned their highest returns of 1.05% per month. The returns started to decline but earned 0.83% per month during the 2008 recession period. For speculative-grade stocks, the returns were 1.68% per month during the 2000s and declined to 1.13% per month during 2008. Therefore, we report that the declining trend of momentum returns is due to the impact of the business cycles.

Next, we contribute in the literature by studying the behaviour of momentum returns for investment-grade and speculative-grade stocks across a business cycle. The reasoning behind this approach is that we know from previous studies which use time series that momentum returns vary with business cycles (see e.g., Avramov and Chordia 2006; Chordia and Shivakumar 2002). Credit risks also vary during business cycles. We use NBER business cycle to observe how the momentum returns of the investment grade stocks and the speculative grade stocks behave across the NBER business cycle. We find that, in a cross section of firms, investment-grade stocks do earn significant momentum returns during both expansion and contraction periods. During an expansion, investment-grade stocks earn 0.80% per month and earn 1.14% per month during contractions. We find that speculative-grade stocks earn as much as 1.20% per month during expansions and 1.61% per month during contractions. We report that the higher momentum returns earned during the contractions are a result of the uncertainty imposed by the business cycle and the uncertainty resulting from the credit ratings of these low credit-rated firms.

We provide a risk-based explanation for momentum returns among different types of credit-rated stocks. Our reasoning is that if the market is efficient, then we expect a risk-based model to explain momentum phenomenon. Like most momentum studies we control for a number of factors. First, we control for the Fama and French (1993) three factors (e.g., Grundy and Martin 2001; Avramov et al. 2007; Jegadeesh and Titman 2001). We report that these factors cannot explain momentum returns either in speculative-grade or in investment-grade stocks. In investment-grade stocks and speculative-grade stocks the alphas are 0.85% and 1.27% respectively and are statistically significant. Next, we control for up and down market states. (e.g., see Cooper et al. 2004; Wang et al. 2009). We show that for credit-rated stocks, the market states cannot explain momentum returns either in investment-grade or in speculative-grade stocks. In speculative-grade stocks, the alpha remains significant and high at 1.24% per month; and for investment-grade stocks, it remains at 0.83% per month. Also, we control for macroeconomic risk factors (e.g., see Chordia and Shivakumar 2002; Avramov and Chordia 2006). We report that macroeconomic risk factors can partially explain momentum returns for both speculative-grade and investment-grade stocks when the market is less volatile. But they do not explain the returns during the tech-boom periods in the early 2000s and subprime financial crisis period in 2008. The empirical results have important insights for researchers and investors; researchers can investigate the behaviour of momentum returns during business cycles while momentum investors can benefit from forming portfolios during market downturn.

The rest of the paper is organized as follows. Section 2.2 presents a literature review. Section 2.3 discusses methodology and data used in the study. Section 2.4 provides the empirical results and Sect. 2.5 concludes.

2.2 Literature Review

This section briefly discusses the studies on momentum returns, momentum among credit-rated stocks, and momentum and common risk factors.

2.2.1 The Persistence of Momentum Returns in Different Dimensions

The literature on momentum returns is highly influenced by the empirical study of Jegadeesh and Titman (1993) who were the first to document that in the US stock market, past winners outperform past losers over 3–12 month periods and earn momentum returns of 12% per annum. Subsequent studies extend the original research in different dimensions including over time (e.g., see Jegadeesh and Titman 1993, 2002), across markets (Chan et al. 2000; Rouwenhorst 1998; Chui et al. 2010) and among different asset classes, such as on currencies (e.g., Okunev and White 2003; Menkhoff et al. 2011), on commodities (e.g., Miffre and Rallis 2007; Gorton et al. 2008), international government bonds (Asness et al. 2012), residential real estate (Beracha and Skiba 2011), and US corporate bonds (Gebhardt et al. 2005; Jostova et al. 2010). Studies also demonstrate that momentum returns are significant among certain subsamples of stocks. For example, Momentum are higher for stocks that are small and low analysts coverage (Hong et al. 2000), high analysts forecast dispersion (Zhang 2006), among large-caps stocks (Obrecht 2006), in firms with high information uncertainty (Jiang et al. 2005; Zhang 2006), among low-credit-rated firms (Avramov et al. 2007) and high turnover (Lee and Swaminathan 2000).

2.2.2 Momentum Returns and Credit Ratings

Avramov et al. (2007) find that momentum payoffs are high in low credit-rated firms and are not observed otherwise. They report that momentum returns are significant in stocks with high credit risk, and this significance remains unexplained when controlling for firm size, firm age, value, turnover, leverage, return volatility, analysts' forecast dispersion, and cash flow volatility. Lee (2012) reports partial confirmation of Avramov et al.'s (2007) results for the US market and finds a reverse trend in the Taiwan market. They report that in Taiwan the highest momentum returns are earned by the high investment-grade group.

Du and Suo (2005) study the behaviour of change in credit ratings on momentum returns and report that the duration effect on the downgrade is a result of the downgrade momentum effect. Blume et al. (1998) study panel data on credit ratings and suggest that the decline in credit ratings is attributable to increasingly stringent standards applied by agencies when deciding the credit quality of corporations. Avramov et al. (2007) demonstrate that credit cycles are crucial in explaining the momentum return of credit-rated stocks. They show that there is a negative relation between credit risk and momentum returns that critically depends on credit cycles. Avramov and Hore (2008) find that momentum interacts with firm-level informational uncertainty measures and credit statuses. Avramov and Hore (2008) report that equilibrium momentum returns concentrate in the interaction between risky cash flows and high credit-risk firms. Momentum returns deteriorate and eventually disappear as leverage or cash flow risk diminishes.

2.2.3 Momentum Returns and Risk Factors

The literature on risk-based explanations shows that momentum returns cannot be explained by Fama and French's three factors (Fama and French 1993; Jegadeesh and Titman 1993; 2001). Momentum literature documents the association of momentum returns to various macroeconomic factors with disputed findings. For example, Chordia and Shivakumar (2002) report that momentum returns can be explained when a set of lagged macroeconomic variables are used. However, Moskowitz and Grinblatt (1999) report that the individual momentum returns in that study mainly come from industry momentum profits. In subsequent studies, Griffin et al. (2003) and Cooper et al. (2004) do not confirm the results of Chordia and Shivakumar (2002).

Chordia and Shivakumar (2002) and Avramov et al. (2007) discover that momentum profits result from the predictability of macroeconomic factors. Antoniou et al. (2007) also show that business-cycle variables and behavioural biases can explain the profitability of momentum trading. Liu and Zhang (2008) indicate that the growth rate of industrial production explains more than half of momentum profits. Bhar and Malliaris (2011) study the changes in fundamental, macroeconomic, and behavioural variables across economic regimes and find that momentum is also highly significant across all three regimes: low, average, and above average volatility. Cooper et al. (2004) report that the risk factor for the market states can explain momentum returns. Defining the two states of the market as up (down) when the lagged 3-year market return is nonnegative (negative), these authors report that short-run momentum profits exclusively follow up periods. In a subsequent study, Hwang and Rubesam (2008) and Lee (2012) confirm the findings of Cooper et al. (2004) and show that momentum disappears when accounted for the market state risk factors.

The above literature demonstrates that the generation of momentum returns from different types of credit-rated stocks and its association with the business cycle is yet to develop. Our study completes this task.

2.3 Data

We use data from all of the stocks listed on the NYSE, AMEX, and NASDAQ from the Centre for Research in Security Prices (CRSP). We use the following selection criteria. Following Jegadeesh and Titman (2001), we include all of the stocks that are priced above $5, have a non-missing observation at the beginning of the holding period, and have at least six consecutive monthly return observations. We include all stocks with S&P ratings in the Compustat database and prices in the CRSP database. Details of S&P ratings are given in the appendix. For S&P credit ratings, we assume that the last quarter rating will continue to the immediately following quarter, until the new rating releases at the end of the quarter. Our research period is from January 1985 to November 2011. For NBER business cycle period we have used the period that covers our sample period from 1985 to 2011. Therefore for NBER business cycle period (see Table 2.4) we have used date from 1982 onward that sufficiently covers our sample period of 1985 and onward. For the screening procedure we apply the following criterion: Some stocks have more than one issue with the same GVKEY. In the GVKEY some of the companies have several stocks issued in the market (all common stocks).[1] While one of them has a rating, the other might not. For our empirical result we have considered the one that does not have rating is as an unrated stock.

Table 2.1 presents the statistics. We start with a total of 15,373 stocks. After the screening process described above, we have a total of 14,665 stocks of which 11,135 are not rated and 3939 are rated. Of the rated stocks, 2054 are investment grade and 2627 are speculative grade.[2] The mean return is 0.05% in the sample. The mean returns non-rated and rated stocks are 0.03% and –0.07% per month respectively. The return volatility for all of the stocks is 12.93%; and, between non-rated and rated stocks, the volatility is 13.22% and 12.58%. Between the two categories of rated stocks, the return volatility is higher among the speculative-grade stocks at 14.95%, while the standard deviation of the investment-grade stocks is 11.03%. The skewness and kourtosis show that the distribution is negatively skewed and fat tailed.

To describe different business cycles, we use the definitions from the National Bureau of Economic Research (NBER). Our sample comprises of four expansion and three contraction periods as defined by NBER. The expansion periods are from November 1982 to July 1990, March 1991 to March 2001, November 2001

[1] Data in the Compustat records the number of issues with and without a rating.

[2] One stock can switch among various grades over its life time, e.g., when a stock is first graded, it can be graded BBB, but later can be graded as B, A, AA, AAA, etc.The numbers given here are an average only.

Table 2.1 Summary statistics. This table presents the summary statistics for the monthly returns across both, the stocks that are not rated and those that are rated by Standard & Poor's (S&P). The sample period is from January 1985 to November 2011. The returns represent the time-series mean of the cross-sectional average return for each month in percentages. The standard deviation, skewness, and the kurtosis are computed as the cross-sectional medians over the stocks in the sample. The *Credit Rated* stocks are those rated by S&P, *Investment Grade* stocks are those rated from Aaa to Bbb (numerical score 1 to 10) and *Speculative Grade* stocks are those rated as Bbb- to C (numerical score from 11 to 21)

Summary statistics					
	All stocks	Not credit rated	Credit rated	Investment grade	Speculative grade
Total no. of stocks after screening (NYSE, NASDAQ, and AMEX)	14,665	11,135	3939	2054	2627
Total no. of months	323	323	323	323	323
Mean (%)	0.05%	0.03%	−0.07%	−0.07%	−0.07%
Standard deviation (%)	12.93%	13.22%	12.58%	11.03%	14.95%
Skewness	−1.20	−0.90	−1.83	−2.84	−0.95
Kourtosis	20.52	17.31	21.92	35.31	11.05
Total no. of stocks before screening (NYSE, NASDAQ, and AMEX)	15,373	11,180	4193	2186	2901
No. of stocks lost due to filtering	708	454	254	132	274

to December 2007, and June 2009 to November 2011. And the three contraction periods are July 1990 to March 1991, March 2001 to November 2001, and December 2007 to June 2009. The Fama-French (1993) three factors comprise the return on the CRSP value-weighted market index in excess of the 1-month Treasury bill rate (*MKT-RF*), the small-minus-big size factor (*SMB*), and the high-minus-low book-to-market ratio factor (*HML*) that we collected from Kenneth French's data library. The definition of the two market states variables is the 36-month lagged average market return (*LAGMKT*) and its square ($LAGMKT^2$) from Cooper et al. (2004). We use the following macroeconomic variables from Chordia and Shivakumar (2002). The dividend yield (*DIV*) is the total dividend payment accrued to the CRSP value-weighted market index over the past 12 months divided by the current price level of the market index. The short rate (*YLD*) is the yield on the 3-month Treasury bill. The term premium (*TERM*) is the yield spread of a 10-year Treasury bond over a 3-month Treasury bill. The default premium (*DEF*) is the yield spread between Moody's Baa and Aaa rated bonds. The data on the macroeconomic variables comes from the Federal Reserve data in the Wharton Research Data Services (WRDS). We conduct a comprehensive and detailed analysis on the behaviour of the momentum returns from credit-rated stocks by first conducting all of our empirical investigations over the entire sample period and then in 5-year subperiods. The subperiod focuses on the impact of the credit risk and the business cycle on the momentum returns.

2.3.1 Methods

First, we calculate the raw momentum returns by following Jegadeesh and Titman (2001). They use 6-month formation (J) and 6-month holding (K) (*JxK = 6x6)* strategies.[3] A month is skipped between the formation and the holding periods. At the end of the holding period, the momentum portfolio becomes the difference between the returns on winner and loser portfolios. We calculate momentum returns for all credit-rated subsamples and NBER business cycles for expansions and contractions. In order to explain the momentum returns, we use three multi-factor regression models with three different types of risk factors. The first is the Fama-French three factors (Fama and French 1993), the second is the market states variables (Cooper et al. 2004), and the third is the macroeconomic variables (Chordia and Shivakumar 2002). We estimate the following model:

$$MR_{CR_{t,6x6}} = \alpha + \sum_{j=1}^{n} \beta_j f_t + \varepsilon_t, \tag{2.1}$$

where $MR_{CR_{t,6x6}}$ is the momentum return generated at time t for the credit-rated category CR (=1, 2, 3) where (1) is *All Rated*, (2) is *Investment Grade*, and (3) is *Speculative Grade*. The f_t (=1, 2, 3) is the vector of risk factors from (1) the Fama-French three factors, (2) the market state variables, and (3) the macroeconomic variables. The β_j (j = 1,…., n) represents the loading for the factors, α is the coefficient estimate for the constant, and ε_t is the residuals with $E(\varepsilon_t) = 0$, $Cov(\varepsilon_t, f_t) = 0$ and ε_t *iid* $(0,\sigma^2)$. Using Eq. (2.1), we test whether the momentum returns of credit-rated stocks remain after accounting for the three different types of risk factors. If the market is efficient and the momentum returns are compensation for risks, then we expect to see that the alpha is equal to zero.

2.4 Empirical Findings

Table 2.2 reports the monthly momentum returns of all stocks regardless of their rating. The first three columns represent the returns on winner and loser portfolios and momentum returns. Next, we calculate the difference between the first two. Column four gives the average decile portfolio size. On average, these stocks

[3]Our sample period is from 1985 to 2011. The first cycle of NBER business cycle that covers our sample period starts from 1982. We use the S&P Long-Term Domestic Issuer Credit Rating. Data on this variable is available on Compustat on a quarterly basis starting from the second quarter of 1985. Again, for stocks to be included in the empirical estimation it has to have at least six-consecutive monthly return observations at the beginning of the holding period. Furthermore, due to the *JxK = 6x6* strategy and also skipping a month, in total $J + K-1$ at least 11 months observation will be lost. Due to these criterions the sub- sample period starts from 1987.

Table 2.2 Momentum returns of all stocks. The following table reports the monthly returns for winner, loser, and momentum portfolios based on the *JxK* = *6x6* strategy (6-month historic returns held for the following 6 months). The sample period is from January 1985 through November 2011. In each month *t* for all NYSE, AMEX, and NASDAQ stocks with returns from $t-6$ through $t-1$ on the monthly CRSP database, the stocks are ranked into decile portfolios according to their returns during the formation period (*J*). The decile portfolios are formed monthly by equally weighting all firms in that decile ranking. Winner and Loser are the equal-weighted portfolios that reflect 10% each of the stocks with the lowest and the highest returns over the previous 6 months respectively. We long the Winner portfolio and short the Loser portfolio and hold the positions for the following holding (*K*) months ($t+1$ *to* $t+6$). The month *t* is skipped between the formation and the holding period. At the end of the holding period, the Momentum portfolio is realized as the difference between the returns on the Winner and the Loser portfolios. Panel A reports the output results for *Not Rated* and panel B reports the output results for *All Rated* stocks. The column *Portfolio size* reports the average size of the decile portfolio during each period. The numbers in bold fonts represent significance at the 5% and 1% levels and the *t*-statistics are given. The table reports the momentum return in percentage, per month, and when excluding penny stocks from the sample. A minimum six-month observation is required for any stocks to be included in the sample.

	All stocks			
Subperiod	Loser	Winner	Momentum	Portfolio Size
1985–2011	−0.60	0.48	**1.08**	413
t-stat	−1.95	2.18	4.36	
1987–1991	−0.61	0.40	**1.01**	279.43
t-stat	*−2.81*	*2.15*	*9.55*	
1992–1996	−0.50	0.53	**1.03**	379.72
t-stat	*−3.78*	*4.28*	*10.45*	
1997–2001	−0.84	0.56	**1.40**	476.37
t-stat	*−2.24*	*2.86*	*4.17*	
2002–2006	−0.48	0.52	**1.00**	456.53
t-stat	*−1.73*	*3.25*	*5.61*	
2007–2011	−0.58	0.40	**0.98**	472.34
t-stat	*−1.68*	*1.19*	*6.56*	

generate statistically significant momentum returns of 1.08% per month during the sample period. All empirical analysis has been conducted on the entire sample period and then further investigation has been made on 5-year sub-sample periods (10-year sub-period). The choice of the sub-periods is based on the consideration of sufficient observations and the availability of the data for the variables used in this study so that meaningful parameter estimates can be obtained.[4] In all of the sub periods the momentum returns are on average 1% per month and 12% per annum. This result is consistent with the findings of Jegadeesh and Titman (1993)

[4]The use of sub-period is common in momentum literature (see among others, Jegadeesh and Titman 2002; Cooper et al. 2004). Studies on historical stock performance can suffer from survivorship bias. Therefore the study of the sub-periods will help mitigate survivorship bias and also help examine if momentum returns and its interaction with risk factors varies in different sub-periods.

who report that momentum returns are on average 1% per month in the US market. We study the two subperiods of 1997–2001 for the tech-boom and the late 2000s financial crisis of 2007–2011. We report that in the subperiod of 1997–2001, the momentum returns are significant and the highest among all other subperiods at 1.40% per month. In the subperiod of 2007–2011, the momentum returns only decline to 0.98% per month and are statistically significant.

Table 2.3 reports the momentum returns for all of the credit-rated stocks and for each category of credit-rated stocks. Panel A of Table 2.3 reports the momentum returns for all of the credit-rated stocks. On average, the momentum returns are 1.03% per month for all of the credit-rated stocks. These stocks earn the highest momentum returns at 1.29% per month in the subperiod of 1997–2001 and 0.99% per month in the subperiod of 2007–2011. Panel B of Table 2.3 reports the momentum returns for the investment-grade stocks. The credit ratings of these stocks are from AAA to BBB. They generate momentum returns of only 0.85% per month throughout the sample period, the lowest among all categories of stocks. In the subperiod of 1997–2001, they earn 1.05% and are significant; and in the subperiod of 2007–2011, they earn 0.83% per month. The low momentum returns for the investment-grade stocks are as expected as these stocks carry low credit risk and impose less uncertainty on the investor. They have low standard deviations as we discussed in Table 2.1. The results are consistent with those reported by (Zhang 2006) that momentum returns provide compensation for high ambiguity (Zhang 2006) and that investment-grade stocks do not impose high ambiguity.

Panel C of Table 2.3 presents the momentum returns from the speculative-grade stocks. The credit ratings of the speculative-grade stocks are from BBB to C. The speculative-grade stocks generate momentum returns of 1.27% per month (15.24%, per annum). This is 42% larger than the returns of the investment-grade stocks. In all of the subperiods, the momentum returns are more than 1%. Notably, in the subperiods of 1997–2001 and 2007–2011, the returns are 1.68% and 1.13% per month respectively. The momentum returns among the speculative-grade stocks are the highest, which implies high uncertainty because of the high default rates in this credit-rating bracket. In Table 2.1, we describe the speculative stocks as having lower average returns and higher standard deviations. Therefore, the high momentum returns of speculative-grade stocks provide compensation for the high uncertainties they impose. The results are consistent with the findings of Avramov et al. (2007) and Lee (2012) who report that momentum returns are high in high credit risk stocks.

In sum, the unrated stocks earn momentum returns of 1.10%, higher than the credit-rated stocks at 1.03%. The results are consistent with those of Avramov et al. (2007) who report that momentum returns for all of the credit-rated stocks are 1.29% per month and for the unrated stocks is 1.43% per month for during the sample period of 1985 through 2003. The difference in the results of this study and Avramov et al. (2007) is our inclusion of the effect from the subprime financial crisis in 2007–2011. The momentum returns of speculative-grade stocks are high, 1.27% per month compared to the investment-grade stocks with 0.85% per month. This difference can be attributed to the uncertainty imposed by the credit ratings.

Table 2.3 Momentum returns of investment-grade and speculative-grade stocks. The following table reports the monthly returns for winner, loser, and momentum portfolios based on $JxK = 6x6$ strategy (six-month historic returns held for the following 6 months). We follow the same procedure as in Table 2.2 for the same sample period, which is from January 1985 through November 2011. In each month t for all stocks with returns from $t - 6$ through $t - 1$ on the monthly CRSP database, the stocks are ranked into decile portfolios according to their returns during the formation period (J). Winner and Loser are the equal-weighted portfolios that reflect 10% each of the stocks with the lowest and the highest returns over the previous 6 months respectively. We long the Winner portfolio and short the Loser portfolio and hold the positions for the following holding (K) months ($t + 1$ *to* $t + 6$). The month t is skipped between the formation and the holding period. At the end of the holding period, the Momentum portfolio is realized as the difference between the returns on the Winner and the Loser portfolios. Panel A reports the output results for *All Rated* stocks, Panel B reports the output for *Investment Grade*, and Panel C reports the output for *Speculative Grade* stocks. The column *Portfolio size* reports the average size of the decile portfolio during each period. The numbers in bold fonts represent significance at the 5% and 1% levels and the t-statistics are given. The table reports the momentum return in percentage, per month, and when excluding penny stocks from the sample. A minimum 6-month observation is required for any stocks to be included in the sample.

	Panel A: all rated				Panel B: investment grade				Panel C: speculative grade			
Subperiod	Loser	Winner	Momentum	Portfolio size	Loser	Winner	Momentum	Portfolio size	Loser	Winner	Momentum	Portfolio size
1985–2011	−0.60	0.43	**1.03**	112.88	−0.52	0.33	**0.85**	74.75	−0.70	0.57	**1.27**	36.56
t-stat	−1.82	1.80	4.19		−1.75	1.67	3.94		−1.77	1.95	3.62	
1987–1991	−0.63	0.34	**0.97**	82.35	−0.55	0.28	**0.83**	57.60	−0.71	0.44	**1.15**	23.37
t-stat	−2.59	1.69	6.92		−2.13	1.56	5.32		−2.98	1.75	5.51	
1992–1996	−0.48	0.44	**0.91**	88.92	−0.43	0.33	**0.76**	64.85	−0.53	0.61	**1.14**	22.88
t-stat	−3.63	3.24	7.46		−2.88	2.78	5.51		−4.04	3.28	7.70	
1997–2001	−0.80	0.50	**1.29**	125.40	−0.66	0.38	**1.05**	86.07	−0.99	0.69	**1.68**	37.57
t-stat	−2.69	3.07	4.07		−3.26	3.11	4.61		−2.37	2.98	4.03	
2002–2006	−0.51	0.49	**0.99**	137.27	−0.42	0.36	**0.78**	87.05	−0.62	0.63	**1.25**	48.42
t-stat	−1.53	2.63	4.85		−1.50	2.53	4.35		−1.41	2.92	3.61	
2007–2011	−0.60	0.39	**0.99**	130.76	−0.55	0.29	**0.83**	78.22	−0.64	0.49	**1.13**	50.80
t-stat	−1.28	1.00	5.14		−1.20	0.88	3.49		−1.38	1.11	5.78	

We next observe the momentum returns from the credit-rated stocks across business cycles and across macroeconomic risk factors. We examine the momentum returns during expansions and contractions as defined by the NBER business cycle. Table 2.4 reports the momentum returns for different credit-rated stocks over the business cycles. Panel A of Table 2.4 reports the momentum returns during expansions for all stocks that are credit rated. On average, the credit-rated stocks generate 0.98% momentum returns per month during the expansions with a range between 1.04% and 0.91% per month. These returns are statistically significant. We find significant momentum returns of 0.80% per month among the investment-grade stocks. In the four expansion periods, the returns range from 0.72% to 0.87% per month and are statistically significant. The momentum returns from the speculative-grade stocks earn on average 1.20% per month and are statistically significant. They earn more than 1% in all of the four expansion periods and range between 1.32% and 1.09% per month. This result is consistent with the findings of Avramov et al. (2007) who find that momentum returns are profitable among firms with high credit risk.

Panel B of Table 2.4 reports the momentum returns during economic contractions. The credit-rated stocks generate momentum returns of 1.36% per month. The momentum returns of the credit-rated stocks range between 1.83% and 1.08% per month during the three contraction periods of July 1990 to March 1991, March 2001 to November 2001, and December 2007 to June 2009. The momentum returns from the investment-grade stocks are 1.14% per month during the contractions. They range between 0.97% and 1.38% per month in the three contraction periods. The momentum returns from the speculative-grade stocks are on average 1.61% per month, which is the highest among the different categories of credit-rated stocks. And in the three contraction periods, they range between 1.23% and 2.35% per month. This result is consistent with the findings of Lee (2012) who finds momentum returns from middle investment-grade stocks during recessions in the US market during the period of 1998 through 2008.

The momentum returns are significant among both investment-grade and speculative-grade stocks during both expansions and contractions. However, the returns are remarkable for speculative-grade stocks during contractions. Uncertainty is higher for speculative-grade stocks that have higher credit risk during periods of contraction when, economy-wide, credit risks increase. The possibility exists that momentum returns are compensation for those risks. If momentum returns are a systematic phenomenon and are just a mere compensation for bearing systematic risk, then the momentum returns should disappear once the appropriate market-wide, common risk factors are accounted for.

Table 2.4 Momentum returns of credit-rated stocks across NBER business cycles. The following table reports the behavior of the monthly returns for winner, loser, and momentum portfolios formed based on $JxK = 6x6$ strategy excluding the penny stocks. The sample period is from January 1985 through November 2011. There are four expansion and three contraction periods as defined by NBER (http://www.nber.org/cycles.html) in the sample period. A minimum 6-month observation is required for a company to be included in the sample. Winner and Loser are the equally weighted portfolios that reflect 10% each of the stocks with the lowest and the highest returns over the previous 6 months respectively. The Winner portfolio is held long and the Loser portfolio is held short for the following holding (K) months ($t + 1$ *to* $t + 6$). The month t is skipped between the formation and the holding period. The Momentum portfolio is the difference between the returns on the Winner and the Loser portfolios. Panel A reports the output for *All Rated* stocks, Panel B and Panel C report the output for *Investment Grade* and *Speculative Grade* stocks respectively. The column *No. of Months* represents the size of each subperiod for different business-cycle periods. The estimates are reported in percentage, the numbers in bold represent significance at the 1% and 5% levels, and the t-statistics are given

Momentum return of credit-rated stocks across NBER business cycle

			Panel A: all rated				Panel B: investment-grade			Panel C: speculative-grade		
	Subperiod	No. of months	t-stat	Loser	Winner	Momentum	Loser	Winner	Momentum	Loser	Winner	Momentum
Panel A: Expansion periods												
Expansion	Nov-1982 – Jul 1990	98		−0.61	0.35	**0.96**	−0.53	0.28	**0.82**	−0.70	0.46	**1.16**
			t-stat	*−2.94*	*2.22*	*8.22*	*−3.40*	*1.98*	*5.06*	*−3.40*	*2.57*	*9.15*
	Mar 1991 – Mar 2001	128		−0.58	0.46	**1.04**	−0.51	0.35	**0.87**	−0.68	0.64	**1.32**
			t-stat	*−2.56*	*2.75*	*4.20*	*−2.60*	*2.64*	*4.33*	*−2.24*	*2.62*	*3.67*
	Nov 2001 – Dec 2007	74		−0.51	0.49	**1.00**	−0.42	0.36	**0.78**	−0.62	0.63	**1.24**
			t-stat	*−1.61*	*2.89*	*4.69*	*−1.58*	*2.78*	*4.25*	*−1.46*	*3.19*	*3.59*
	Jun 2009 – Nov 2011	32		−0.34	0.57	**0.91**	−0.29	0.43	**0.72**	−0.40	0.69	**1.09**
			t-stat	*−1.00*	*1.71*	*5.16*	*−1.02*	*1.41*	*3.86*	*−1.04*	*1.85*	*5.38*
	Average			−0.51	0.47	**0.98**	−0.44	0.36	**0.80**	−0.60	0.60	**1.20**
			t-stat	*−1.87*	*2.26*	*5.19*	*−1.78*	*2.01*	*4.35*	*−1.82*	*2.43*	*4.65*

(continued)

Table 2.4 (continued)

Momentum return of credit-rated stocks across NBER business cycle

			Panel A: all rated				Panel B: investment-grade			Panel C: speculative-grade		
	Subperiod	No. of months	t-stat	Loser	Winner	Momentum	Loser	Winner	Momentum	Loser	Winner	Momentum
Panel B: Contraction periods												
Contraction	Jul 1990 – Mar 1991	9		−0.93	0.15	**1.08**	−0.86	0.11	**0.97**	−0.99	0.24	**1.23**
			t-stat	*−4.57*	*1.16*	*12.20*	*−4.51*	*0.91*	*12.98*	*−3.90*	*1.37*	*9.79*
	Mar 2001 – Nov 2001	9		−1.26	0.57	**1.83**	−0.88	0.50	**1.38**	−1.66	0.69	**2.35**
			t-stat	*−6.06*	*6.24*	*9.76*	*−5.91*	*5.53*	*10.83*	*−6.08*	*6.17*	*8.92*
	Dec 2007 – Jun 2009	19		−1.14	0.02	**1.17**	−1.09	−0.01	**1.08**	−1.17	0.08	**1.25**
			t-stat	*−3.77*	*0.06*	*7.90*	*−3.57*	*−0.03*	*7.16*	*−4.15*	*0.19*	*6.54*
	Average			−1.11	0.25	**1.36**	−0.94	0.20	**1.14**	−1.27	0.34	**1.61**
			t-stat	*−4.66*	*1.30*	*9.62*	*−4.39*	*1.26*	*9.71*	*−4.72*	*1.44*	*8.32*

2.4.1 Can the Fama-French Three Factors Explain Momentum Returns in Credit-Rated Stocks?

In this subsection, we test whether the momentum returns disappear in credit-rated stocks when we take into account the Fama-French three factors. Because the literature reports that momentum in general cannot be explained by the Fama-French three factors, we investigate if this conclusion holds for credit-rated stocks. Table 2.5 presents the coefficient estimates for Eq. (2.1) where f is a vector of the Fama-French three factors. Panel A reports the results for all of the credit-rated stocks. The momentum returns are significant at 1.03% per month for all of the credit-rated stocks after controlling for the Fama-French three factors. The alpha for all of the credit-rated stocks is statistically significant in the whole sample period and in all of the other subsample periods. Panel B reports the results for the investment-grade stocks when the Fama-French three factors are accounted for. For the investment-grade stocks, the momentum returns are 0.85% per month after controlling for these three risk factors. They are also significant at over 0.75% per month in all of the subperiods. Panel C reports the results for the speculative-grade stocks. The alpha is 1.27% per month and is statistically significant. In all of the subperiods, the alphas are significant and range between 1.102% and 1.69% per month. Our results show that the momentum returns of the credit-rated stocks remain unexplained when the Fama-French three factors are accounted for. The findings confirm the results of various earlier studies in the momentum literature (e.g., Avramov et al. 2007; Fama and French 1996; Grundy and Martin 2001).

2.4.2 Can Market States Explain the Momentum Returns in Credit-Rated Stocks?

Table 2.6 reports the coefficient estimates for Eq. (2.1) where f is the vector of the 36-month lagged average market return (*LAGMKT*) and its square ($LAGMKT^2$). Panel A reports the results for all of the credit-rated stocks. The alpha is significant at 1.02% per month. Also in different subperiods, we observe significant alphas ranging between 1% and 2.02% per month. Panel B reports the results for the investment-grade stocks. The alpha for the investment-grade stocks is on average 0.83% per month and ranges from 0.78% to 1.48% per month in different subperiods. In the subperiods of 1997 to 2001 and 2007 to 2011, the alphas are 1.48% and 0.84% per month respectively. Panel C reports the momentum returns for the speculative-grade stocks. They remain significant in all of the subperiods and range between 0.86% and 2.63% per month. In the subperiods of 1997 to 2001 and 2007 to 2011, the alphas are 2.63% and 1.16% per month respectively. The results clearly depict that the momentum returns from credit-rated stocks remain significant after controlling for the market's states. These results are inconsistent with those reported by Cooper et al. (2004) and Wang et al. (2009) who report that market states variables can explain the momentum returns in stocks that are traded in the US market and in the Taiwan market.

Table 2.5 Momentum returns of credit-rated stocks and Fama French three factors. The Winner, Loser, and Momentum portfolios are formed based on the strategy described in Table 2.2 with $JxK = 6x6$ and excluding penny stocks. The following table represents the coefficients and the *t*-statistics obtained when the momentum returns of each type of credit-rated stock, *All Rated*, *Investment Grade*, and *Speculative Grade*, is regressed on the Fama-French three factor variables: *MKT_ RF*, *SMB*, and *HML*. The *MKT_ RF* is the monthly return on the CRSP value-weighted market index in excess of the 6-month Treasury bill rate. The *SMB* and *HML* are the Small-Minus-Big size factor and the High-Minus-Low book-to-market ratio factor respectively. The regression model comprises X as the vector of the Fama-French factors. The regression is carried out separately for $MR_{t,6x6} = \alpha + \sum_{j=1}^{n} \beta_j f_t + \varepsilon_t$ for each subperiod. The coefficient covariance of the regression is derived from White's heteroskedasticity consistent coefficient covariance. The numbers are reported in percentages and numbers in bold represent significance at the 5 and 1% levels. The *t*-statistics and the adjusted *R*-squared are also given

Panel A: all rated						Panel B: investment grade					Panel C: speculative grade				
					Adj					Adj					Adj
Period	Alpha	Mkt_Rf	SMB	HML	R-Squared	Alpha	Mkt_Rf	SMB	HML	R-Squared	Alpha	Mkt_Rf	SMB	HML	R-Squared
1987–2011	**1.03**	−0.004	0.006	0.013	2.25%	**0.851**	−0.004	0.004	0.009	1.64%	**1.267**	−0.003	0.006	**0.02**	2.47%
t-stat	*72*	*−1.169*	*1.22*	*2.62*		*67.54*	*−1.51*	*0.96*	*2.034*		*62.29*	*−0.869*	*1.03*	*2.89*	
1987–1991	**0.97**	0.001	0	0.006	−4.71%	**0.836**	0.005	0.005	0.012	−2.19%	**1.151**	−0.003	−0.01	−0.01	−4.13%
t-stat	*51.2*	*0.335*	*−0*	*0.59*		*39.84*	*0.939*	*0.57*	*1.021*		*40.694*	*−0.444*	*−0.4*	*−0.7*	
1992–1996	**0.89**	0.002	0.011	0.026	19.93%	**0.749**	−0.003	0.012	**0.026**	16.65%	**1.102**	0.011	**0.02**	**0.03**	24.19%
t-stat	*55.5*	*0.287*	*1.88*	*4.02*		*40.3*	*−0.368*	*1.68*	*3.479*		*58.423*	*1.54*	*2.23*	*4.59*	
1997–2001	**1.3**	−0.008	0.001	−0	−3.77%	**1.055**	−0.008	−0.01	−0.005	−2.23%	**1.686**	−0.009	0.007	0.005	−2.66%
t-stat	*29.8*	*−0.765*	*0.08*	*−0.1*		*34.1*	*−1.165*	*−0.7*	*−0.606*		*29.66*	*−0.68*	*0.52*	*0.31*	
2002–2006	**0.99**	**−0.019**	0.011	0.005	6.21%	**0.784**	−0.019	0.013	−0.002	8.17%	**1.248**	**−0.028**	0.012	0.015	4.63%
t-stat	*35.2*	*−2.383*	*1.01*	*0.38*		*32.13*	*−2.751*	*1.4*	*−0.151*		*25.928*	*−2.064*	*0.67*	*0.64*	
2007–2011	**0.989**	**−0.012**	0.03	0.023	22.57%	**0.828**	**−0.016**	**0.04**	0.018	14.58%	**1.138**	**−0.01**	**0.02**	**0.03**	21.52%
t-stat	*44*	*−2.725*	*2.9*	*2.64*		*28.33*	*−2.661*	*2.64*	*1.574*		*49.245*	*−2.154*	*2.1*	*3.25*	

Table 2.6 Momentum returns of credit-rated stocks and market states. The following table represents the coefficients and the t-statistics obtained when momentum returns of each type of credit-rated stock, *All Rated*, *Investment Grade*, and *Speculative Grade*, is regressed on the market state variables of *LGMKT* and *LGMKT2* as in Cooper et al. (2004). The *LGMKT* and *LGMKT2* are defined as the lagged 36-month market return and the square of the lagged 36-month market return respectively. The returns on the Winner, Loser, and Momentum portfolios are formed based on the strategy described in Table 2.2 for the $JxK = 6x6$ strategy and excluding the penny stocks. The regression model is $MR_{t,6x6} = \alpha + \sum_{j=1}^{n} \beta_j f_{t-1} + \varepsilon_t$ in which X is the vector of the two market state variables. The regression is carried out separately for each subperiod. The coefficient covariance of the regression is derived from White's heteroskedasticity consistent coefficient covariance. The numbers are reported in percentages and the numbers in bold represent significance at the 1% and 5% levels. The t-statistics and adjusted R-squared are also given. Panel A reports the output for *All Rated* stocks while Panels B and C represent the results of the Investment *Grade* and *Speculative Grade* rated stocks respectively

Panel A: all rated					Panel B: investment grade				Panel C: speculative grade			
Period	Alpha	LGMKT	LGMKTSQR	Adj R-Squared	Alpha	LGMKT	LGMKTSQR	Adj R-Squared	Alpha	LGMKT	LGMKTSQR	Adj R-Squared
1987–2011	**1.02**	**−6.71**	**513.05**	2.08%	**0.83**	**−8.63**	**669.32**	5.41%	**1.24**	−5.87	**615.06**	1.47%
t-stat	46.99	−2.35	2.89		44.52	−3.51	4.36		39.94	−1.44	2.42	
1987–1991	**1.21**	**−44.93**	**1844.52**	13.85%	**1.04**	**−42.32**	**1833.51**	12.15%	**1.41**	−41.2	1481.82	20.81%
t-stat	11.39	−2.83	3.2		8.68	−2.35	2.81		8.29	−1.62	1.6	
1992–1996	**1.01**	−27.06	**1619.89**	0.24%	**1.13**	−71.72	3351.95	−0.35%	**0.86**	27.55	−166.75	12.51%
t-stat	3.82	−0.55	0.73		3.77	−1.28	1.32		2.88	0.5	−0.07	
1997–2001	**2.02**	**−72.18**	1460.85	42.97%	**1.48**	**−40.86**	757.9	30.57%	**2.63**	**−95.82**	1973.39	42.57%
t-stat	13.86	−2.83	1.49		12.87	−2.03	0.98		13.74	−2.85	1.53	
2002–2006	**1.03**	**−9.58**	−99.92	14.15%	**0.78**	**−12.19**	408.89	24.33%	**1.34**	**−11.87**	−782.83	12.76%
t-stat	28.53	−2.89	−0.26		26.43	−4.49	1.31		21.89	−2.1	−1.21	
2007–2011	**1**	−3.29	−27.83	−2.17%	**0.84**	**−11.58**	414.87	4.97%	**1.16**	2.58	−730.53	−1.13%
t-stat	27.55	−0.73	−0.04		19.38	−2.16	0.56		31.63	0.56	−1.16	

2.4.3 Can Macroeconomic Factors Explain the Momentum Returns in Credit-Rated Stocks?

Table 2.7 reports the coefficient estimates for Eq. (2.1) when f is the vector of the lagged macroeconomic variables. Panel A of Table 2.7 reports the results for all of the credit-rated stocks. We report interesting findings for the momentum returns from the credit-rated stocks when accounting for the macroeconomic variables. For all credit-rated stocks, the alpha is significant during the sample period of 1987–2011 at 0.96% per month. However, they are significant in only two subperiods: 1992–1996 and 1997–2001 at 0.85% and 2.21% per month respectively. The r-squared for the subperiod of 1997–2011 is 38.94% that indicates the explanatory power of the variables. This finding implies that macroeconomic variables can partially explain the momentum returns and those market-wide macroeconomic variables cannot explain momentum returns when the market is volatile.

Panel B reports the coefficient estimates for the investment-grade stocks. The coefficient estimate for alpha is statistically significant during the entire sample period at 0.59% per month. Among the different subperiods, it is statistically significant in three: 1992–1996, 1997–2001, and 2007–2011 at 0.83%, 1.58%, and 0.83% respectively. Interestingly, the momentum returns are not explained by the macroeconomic variables in the two most important crisis periods: 1997–2001 and 2007–2011. The alpha is not significant in the subperiods of 1987–1991 and 2002–2006 when the US market was less volatile. This finding implies that, during periods of high market volatility, the market-wide macroeconomic variables cannot capture the effect of the momentum phenomenon and that the momentum returns from the investment-grade stocks is a compensation for the risk or uncertainty imposed by the business cycle.

Panel C reports the momentum returns for speculative-grade stocks. The alpha for the speculative-grade stocks is, on average, 1.35% that is statistically significant. In the different subperiods, the alpha is significant for the periods of 1992–1996, 1997–2001, and 2002–2006 at 0.87%, 3.11%, and 0.83% respectively. The alpha is not significant in the subperiods of 1987–1991 and 2007–2011. We estimate the momentum returns with f being a vector of the contemporaneous values of the same macroeconomic variables in Eq. (2.1), and our conclusions do not change.[5] In sum, the results from Table 2.7 show that the momentum returns from credit-rated stocks are partly explained by the macroeconomic variables. The t-statistics shows that the two variables TERM and YLD have greater impact on the speculativegrade

[5]Results are available from authors upon request.

Table 2.7 Momentum returns of credit-rated stocks and macroeconomic risk factors. This table represents the coefficients and the *t*-statistics obtained from the regression in which the momentum returns of each type of credit-rated stock, *All Rated*, *Investment Grade*, and *Speculative Grade*, is regressed over the lagged macroeconomic variables. Momentum returns are formed based on the strategy described in Table 2.2 with a strategy of $JxK = 6x6$ excluding penny stocks. The macroeconomic variables are the dividend yield (*DIV*), the short rate (*YLD*), the term premium (*TERM*), and the default premium (*DEF*). The *DIV* is defined as the total dividend payment accrued to the CRSP value-weighted market index over the past 12 months divided by the current price level of the index. The *YLD* is the yield on the 3 month Treasury bill. The *TERM* is defined as the yield spread of a 10-year Treasury bond and a 3-month Treasury bill. The *DEF* is the yield spread of Moody's Baa and Aaa rated bonds. The regression model is $MR_{t,6x6} = \alpha + \sum_{j=1}^{n} \beta_j f_{t-1} + \varepsilon_t$ in which X is the vector of the lagged macroeconomic variables. The coefficient covariance of the regression is derived from White's heteroskedasticity consistent coefficient covariance. The numbers are reported in percentages and the numbers in bold represent significance at the 1% and 5% levels respectively. The *t*-statistics and adjusted *R*-squared are also given. Panels A, B, and C represent the results of *All Rated*, *Investment Grade*, and *Speculative Grade* stocks respectively

Panel A: all rated							Panel B: investment grade						Panel C: speculative grade					
Period	Alpha	DEF	DIV	TERM	YLD	RSquared	Alpha	DEF	DIV	TERM	YLD	RSquared	Alpha	DEF	DIV	TERM	YLD	RSquared
1987–2011	**0.96**	**0.2**	**−125.4**	0.02	**0.03**	12.67%	**0.59**	**0.23**	**−104.3**	**0.04**	**0.04**	17.71%	**1.35**	0.1	−182.9	0.02	0.03	12.81%
t-stat	14.5	4.28	−6.33	1.44	2.88		10.51	7.19	−6.18	3.25	4.69		14.34	**1.87**	−6.5	0.79	2.5	
1987–1991	0.35	**0.4**	1.5	0.05	0.02	24.01%	0.18	**0.54**	0.19	0.05	0	48.44%	**0.86**	0.15	9.87	0.03	0.01	−3.91%
t-stat	1.32	3.15	0.09	1.5	0.72		0.76	4.88	0.01	1.62	0.01		1.86	0.72	0.32	0.48	0.18	
1992–1996	**0.85**	−0	−10.02	**0.05**	−0.01	15.59%	**0.83**	−0.31	−6.59	**0.09**	−0.01	35.68%	**0.87**	**0.6**	−0.2	−0.02	−0.02	4.58%
t-stat	4.27	−0.1	−0.39	1.98	−0.25		4.22	−1.48	−0.26	3.52	−0.46		3.4	2.2	−0.01	−0.85	−0.67	
1997–2001	**2.21**	0.43	**−381.3**	−0.03	**−0.2**	38.94%	**1.58**	0.4	**−254.6**	−0.03	**−0.1**	37.42%	**3.11**	0.49	−511.6	−0.09	**−0.2**	38.07%
t-stat	3.92	1.47	−3.06	−0.4	−2.41		**3.87**	1.89	−2.82	−0.66	−2.19		4.17	1.26	−3.11	−0.88	−2.68	
2002–2006	−0.13	**0.6**	27.16	**0.11**	**0.1**	50.84%	−0.07	**0.64**	10.63	0.05	0.04	62.50%	**−0.83**	**1**	52.9	**0.26**	**0.21**	49.68%
t-stat	−0.52	6.67	0.61	2.17	2.07		−0.37	8.76	0.31	1.33	1.19		−1.97	6.13	0.69	2.85	2.69	
2007–2011	0.01	**0.1**	3.41	**0.25**	**0.18**	38.61%	**0.83**	**−0**	**0.04**	0.02	**0.83**	14.57%	**0.37**	0.06	−38.73	**0.24**	**0.16**	24.27%
t-stat	0.04	3.98	0.08	5.13	5.25		28.33	−2.66	2.64	1.57	28.3		1.77	1.62	−0.81	4.37	4.17	

momentum returns during different sub-periods. This finding is consistent with the study of Chordia and Shivakumar (2002) who repost that momentum returns remain unexplained when using macroeconomic variables and that the results are strong for some variables particularly TERM and YLD. The momentum phenomenon is more affected by the credit risk and the risk imposed by the uncertainty associated during a market downturn.

2.5 Conclusions

In this paper, we demonstrate that momentum returns are robust in credit-rated stocks and across business cycles. They are remarkable among the speculative-grade stocks during contractions. The momentum returns from speculative-grade and investment-grade stocks are explained when controlled for macroeconomic risk factors and are more prevalent during market upturns and when the market is less volatile. They remain unexplained by the macroeconomic factors during market downturns. The momentum returns provide compensation for the uncertainties imposed on investors because of high credit risk in individual companies and the uncertainties imposed by the business cycle. Zhang (2006) measures ambiguity as the arrival of public information and shows that the momentum profits in high information, ambiguous stocks remain unexplained. Avramov et al.'s (2007) results imply that momentum returns are higher during recessionary periods when credit risk is high. We show that this is indeed the case. Momentum is because of the uncertainty disclosed by the credit risk and the uncertainty imposed by the business cycle. Studies have emphasized on the impact of investor's behavior on momentum returns during business cycle particularly during market downturn. Future study can address the association between investors' behavior and momentum returns during market downturn.

Appendix: S&P Credit Rating

Following Avramov et al. (2007), we use the S&P Long-Term Domestic Issuer Credit Rating that is available from Compustat on a quarterly basis starting with the second quarter of 1985. The S&P rating are as follows:

Credit ratings	Numerical value
AAA	1
Aa+	2
AA	3
Aa−	4
A+	5
A	6
A−	7
BBB+	8
BBB	9
BBB−	10
Bb+	11
BB	12
Bb−	13
B+	14
B	15
B−	16
CCC+	17
CCC	18
CCC−	19
CC	20
C	21
D	22

References

A. Antoniou, H.Y.T. Lam, K. Paudyal, Profitability of momentum strategies in international markets: the role of business cycle variables and behavioral biases. J Bank Financ. **31**, 955–972 (2007)

Asness, C.S., Moskowitz, T.J., Pedersen, L.H., Value and momentum everywhere Chicago Booth Research Paper No. 12–53 (2012)

D. Avramov, T. Chordia, Asset pricing models and financial market anomalies. Rev. Financ. Stud. **19**, 1001–1040 (2006)

D. Avramov, T. Chordia, G. Jostova, A. Philipov, Momentum and credit rating. J. Financ. **62**(5), 2503–2520 (2007)

Avramov, D., Hore, S., Momentum, information uncertainty and leverage—an explanation based on recursive preference. Working Paper, SSRN. (2008), http://papers.ssrn.com/abstract=1017419

E. Beracha, H. Skiba, Momentum in residential real estate. J. Real Estate Financ. Econ **43**(3), 299–320 (2011)

R. Bhar, A.G. Malliaris, Dividends, momentum, and macroeconomic variables as determinants of the US equity premium across economic regimes. Rev. Behav. Financ. **3**(1), 27–53 (2011)

Bhattacharya, D., Kumar, R., Sonaer, G., Momentum loses its momentum: implications for market efficiency. Working Paper, SSRN. (2011), http://papers.ssrn.com/abstract=1928764

M.E. Blume, F. Lim, A.C. MacKinlay, The declining credit quality of U.S. corporate debt: Myth or reality? J. Financ. **53**, 1389–1413 (1998)

K. Chan, A. Hameed, W. Tong, Profitability of momentum strategies in international equity markets. J. Financ. Quant. Anal. **35**, 153–172 (2000)
T. Chordia, L. Shivakumar, Momentum, business cycle and time-varying expected returns. J. Financ. **57**, 985–1019 (2002)
A.C.W. Chui, S. Titman, K.J. Wei, Individualism and momentum around the world. J. Financ. **65**, 361–392 (2010)
M.J. Cooper, R.C. Gutierrez Jr., A. Hameed, Market states and momentum. J. Financ. **59**(3), 1345–1365 (2004)
Du, Y., Suo, W., An empirical study on credit rating change behavior. Working Paper, SSRN. (2005), http://papers.ssrn.com/abstract=670325
A. Eisdorfer, Delisted firms and momentum profits. J. Financ. Markets **11**, 160–179 (2008)
E. Fama, K. French, Common risk factors in the returns on stocks and bonds. J. Financ. Econ. **33**, 3–56 (1993)
E. Fama, K. French, Multifactor explanations of asset pricing anomalies. J. Financ. **51**, 55–84 (1996)
W.R. Gebhardt, S. Hvidkjaer, B. Swaminathan, Stock and bond market interaction does momentum spill over? J. Financ. Econ. **75**, 651–690 (2005)
Gorton, G.B., Hayashi, F., Rouwenhorst, G.K., The fundamentals of commodity futures returns. Yale ICF Working Paper No. 07–08 (2008)
J.M. Griffin, S. Ji, J.S. Martin, Momentum investing and business cycle risk: evidence from pole to pole. J Finance **53**, 2515–2547 (2003)
B.D. Grundy, J.S. Martin, Understanding the nature of risks and the sources of rewards to momentum investing. Rev. Financ. Stud. **14**, 29–78 (2001)
C.R. Harvey, A. Siddique, Stausal skewness in asset pricing tests. J. Financ. **55**, 1263–1295 (2000)
Hwang, S., Rubesam, A., The disappearance of momentum. Working Paper, SSRN. (2008), http://papers.ssrn.com/abstract 968176
H. Hong, T. Lim, J.C. Stein, Bad news travels slowly: size, analyst coverage and the profitability of momentum strategies. J. Financ. **55**, 265–295 (2000)
N. Jegadeesh, S. Titman, Returns to buying winners and selling losers: implications for stock market efficiency. J. Financ. **48**, 65–91 (1993)
N. Jegadeesh, S. Titman, Profitability of momentum strategies: an evaluation of alternative explanations. J. Financ. **56**, 699–720 (2001)
N. Jegadeesh, S. Titman, Cross-sectional and time-series determinants of momentum returns. Rev. Financ. Stud **15**, 143–157 (2002)
G. Jiang, C.M.C. Lee, Y. Zhang, Information uncertainty and expected returns. Rev. Acc. Stud. **10**, 185–221 (2005)
Jostova, G., Nikolova, S., Philipov, A., Stahel, C., Momentum in corporate bond re-turns. Working Paper, George Washington University, 2010
R. Korajczyk, R. Sadka, Are momentum profits robust to trading costs? J Financ. **59**, 1039–1082 (2004)
C.M.C. Lee, B. Swaminathan, Price momentum and trading volume. J Financ. **55**(5), 2017–2069 (2000)
N. Lee, Firm ratings, momentum investment strategies, and market crisis: evidence from US and Taiwan market. Fin Mkt and Port Mgt **26**(4), 449–468 (2012)
L. Liu, L. Zhang, Momentum profits, factor pricing, and macroeconomic risk. Rev. Financ. Stud. **21**, 2417–2448 (2008)
L. Menkhoff, L. Sarno, M. Schmeling, A. Schrimpf, Carry trades and global foreign exchange volatility. J. Financ. **67**(2), 681–718 (2011)
Miffre J., Rallis, G., Momentum strategies in commodity futures markets. EDHEC Risk and Asset Management Research Centre, Working Paper. (2007)
T. Moskowitz, M. Grinblatt, Do industries explain momentum? J Financ. **54**, 1249–1290 (1999)
Obrecht, G. D., Alpha generating momentum strategies. Working Paper, SSRN. (2006), http://ssrn.com/abstract=1180660
J. Okunev, D. White, Do momentum based strategies still work in foreign currency markets? J. Financ. Quant. Anal. **38**, 425–447 (2003)

K. Rouwenhorst, International momentum strategies. J. Financ. **53**, 267–284 (1998)
K. Wang, C. Jiang, Y. Huang, Market states and the profitability of momentum strategies: evidence from the Taiwan stock exchange. Inter. J. Bus. Financ. Res. **3**(1), 89–102 (2009)
Wang, K.Q., Xu, J., Market volatility and momentum. Working paper, ssrn. (2009), http://papers.ssrn.com/abstract=1342719
X.F. Zhang, Information uncertainty and stock returns. J. Financ. **61**(1), 105–136 (2006)

Chapter 3
Drivers of LBO Operating Performance: An Empirical Investigation in Asia

Aurélie Sannajust and Alain Chevalier

Abstract We extend our worldwide research on private equity by studying the drivers of LBO operating performance in Asia. We consider a large set of candidate drivers (financial, governance, macroeconomic, cultural, microeconomics and industry variables) and study their effects on performance over the short- and long-terms. To conduct our study, we use Capital IQ as a data base as well as a hand collected dataset covering LBOs in Asia. We contribute to the current literature by doing an investigation of the impact of macroeconomic factors on the performance of LBOs in Asia. We use a sample of 156 LBO transactions which occurred between 2000 and 2010. Our results show that GDP growth, industry growth, and market return are important drivers that significantly contribute to create value in LBOs.

Keywords Asia • Going private • Delisting • Drivers • Central state ownership • Macroeconomic variables

JEL Classification G24, G34

3.1 Introduction

After the inception of the private equity in the US, this activity spread to Europe and then to many other countries. Investors were seeking for high returns and motivated by the attractive economic growth rates. Hence, increasing investment capital flows into Asian Private Equity occured. According to the Emerging Markets Private Equity Association, the number of leveraged finance transactions completed in emerging Asia in the first 3 months of 2014 rose by 28 percent to 169 versus the same period of 2013. Olivier Carcy, Head of private equity at Credit Agricole's

A. Sannajust (✉)
Finance Department, COACTIS, 42023 Saint-Etienne Cedex 2, Saint Etienne, France
e-mail: aurelie.sannajust@univ-st-etienne.fr

A. Chevalier
Finance Department-ESCP Europe, 75 553 PARIS Cedex 11, Paris, France
e-mail: chevalier@escpeurope.eu

G. Consigli et al. (eds.), *Handbook of Recent Advances in Commodity and Financial Modeling*, International Series in Operations Research & Management Science 257,
DOI 10.1007/978-3-319-61320-8_3

private banking unit, said "in Asia, the leveraged loan market will emerge quite significantly over the next few years because companies have reached a certain maturity over the last 15 years; entrepreneurs want to cash in and that will lead to more transactions".

The internationalization of LBO activities was also a result of the transmission of investment practices and financial technologies from developed markets to emerging Asian markets, Cumming et al. (2010). The development of the LBO market in Asia took place in the 1990's when private equity firms raised dedicated pools of capital in order to maintain and/or increase the number of control investments in Asian private companies.

Actually, there are structural differences between private equity fundraising in Asian countries, as well as in many other emerging markets and developed countries, Oberli (2014). Unlike developed economies where the landscape is formed by large players, Asian markets are structured around sets of fragmented industries with a lot of small companies offering high growth rates. Hence, Asian private equity has been driven during the last 10 years by growth financing constraints for small and mid-sized companies, Naqi and Samanthala (2007). Seeking alternative financial solutions was motivated by the problems of credit availability and financial squeeze for small business, especially after the Asian currency crisis. This situation leads to an increasing public concern and awareness on small business, Berger and Udell, (1998).

Regarding operational changes, Asian LBOs involve acquisitions more than divestitures, Cumming et al. (2010).

The aim of this article is to identify all the drivers that affect the performance of a going private transaction. Our paper contributes to the literature on Private Equity operations because we study performance of going private with financial, capital structure, macro economic, industry level, employment drivers and the influence of states. A large number of the available papers only focus on financial and management factors. While several previous studies have examined the effect of macroeconomic factors on fund-level returns, Phalippou and Zollo (2005), Ljungqvist et al. (2008), Cumming and Walz (2010), Diller and Kaserer (2009), the impact of macroeconomic and industry factors on portfolio firm level returns largely remains unclear. Nikoskelainen and Wright (2007) extend the analysis to include the effects of governance variables but employ a less-developed methodology and do not examine macroeconomic factors. Guo et al. (2011) study the impact of operational improvements and changes in market valuations on investment-level returns. Nevertheless, they exclude the majority of important macroeconomic variables and use a small size sample with limited adjustments for selection bias. Thus, we think they do not reproduce an accurate picture of the drivers of holding period returns in buyouts at the firm level taking into account the impact of firm-level, industry-level and macroeconomic-level factors. Our results are based on analyses with multiple sources of collected data of LBOs in Asia. We use Capital IQ as main database together with a hand-collected dataset for some missed data consulting websites of firms. We select all LBOs between 2000 and 2010. We examine both the drivers of operating performance of LBOs before the going private announcement and the post-delisting performance. The post-delisting performance is usually neglected in the literature due to the difficulties to obtain reliable data on private companies.

The results show that GDP growth, industry growth, and market return are important drivers that significantly contribute to value creation in LBOs. Macroeconomic variables have an important impact on LBOs in Asia especially as the country growth rate is high. Governance variables present important results. These results can be justified by the presence of information asymmetries that cause insider-driven MBOs to benefit from a general uplift in their sector or the economy because management has identified opportunities to exploit or reinvigorate, Wright et al. (2000). Ownership in Asia is dominated by the large presence of governance (Pessarossi and Weill 2013). It shows that governance (central and local) have negative and significant results. They have a very important power in the firm's decision. Indeed, leverage has a positive impact on LBO performance because buyout investors efficiently use debt to improve the equity returns of successful transactions. Moreover, we analyze the differences between LBOs and non-LBOs. We find that LBOs targets in the sample present higher operating performances than non-LBOs. Non-LBOs are smaller than LBOs according to total assets. The level of free cash-flow is also a key factor. Finally, we examine the impact of going private transactions on employees. Going private transactions are also a way to restructure the workforce by adjusting the number of employees. This has a positive impact on the firm's performance by increasing the profit per employee. This interpretation is validated 1 year period after the transaction but we notice an increase in the number of employees 3 years after.

Our study includes five main contributions to the literature. First, we design a LBO sample devoted to emerging markets and especially to Asia. To the best of our knowledge, it is the first time an empirical academic article is written on LBOs in Asia with a long term horizon. Absence of empirical studies is explained by the fact that data are difficult to obtain even if Asia is an interesting region for researchers working on LBOs due to its attractiveness in terms of growth of private equity funds. Second, to understand the value creation process in LBOs, we divided our sample in three regions with three dummy variables to avoid a geographical clustering. Third, we do not focus on the moment where the transaction is signed like many studies do but on the consequences before and after the delisting. We consider on the one hand the operating performance between year -1 and year $+1$ and on the other hand between year -1 and year $+3$. Generally, only market reactions around the acquisition announcement are analyzed by previous contributions. Post-performance is not considered due to the lack of data. Fourth, we take into account the macroeconomic effects on performance of LBOs. It is the first research dealing with the impact of macroeconomic factors on performance of LBOs in Asia. Fifth, we study the presence of ownership in Asia with two variables coming from the Center of China Security Index database. They are important factors because states and other communities have large ownerships in Asian countries. Finally, we analyze the impact of going private decisions on employees.

The remainder of the paper is organized as follows. In Sect. 3.2, we present a literature review and formulate the hypotheses to be tested empirically. Section 3.3 presents the institutional background of emerging economies. Section 3.4 provides a dataset picture and introduces to the reader the main descriptive statistics. Section 3.5 studies the change in operating performances before and after the going private transactions and the impact of going private decisions on employees. Section 3.6 summarizes the main results and brings a few questions for future research.

3.2 Literature Review

To understand the main drivers of operating performance in LBOs, we build on the recent literature that suggests that both buyout deals characteristics (financial variables, capital structure variables), macroeconomic and industry influences are co-determinants of value creation.

Agency costs are incurred by shareholders, or principals, as a result of the separation of ownership and control, Jensen and Meckling (1976). Information asymmetry means that managers are able to pursue objectives such as corporate size rather than maximizing shareholder wealth. A number of corporate governance mechanisms may be used to reduce the extent of the agency costs incurred by the principals. There are two main categories of governance mechanisms, internal and external. Internal mechanisms can be split into monitoring and incentive related. Monitoring mechanisms refer to board structures, Fama (1980), Fama and Jensen (1983), Cadbury (1992), Greenbury (1995), external shareholdings Shivdasani (1993), and debt, Jensen (1986b). The key incentive mechanism is internal shareholdings, Jensen and Meckling (1976). The main external corporate governance mechanism is the market for corporate control, Manne (1965), Jensen (1986b), which acts as the mechanism of last resort if the internal mechanisms fail.

Then the issue becomes how to explain that PTPs (public to private) reduce agency costs. The literature on public-to-private transactions can be split into a number of strands. In this paper, we focus on five main strands.

3.2.1 Tax Benefit

Going private transactions imply an increase in leverage. This leads to an important deduction of interests which is a main source of expected wealth gains. It is a major tax shield increasing the pre-recapitalization value. However, it depends of the fiscal regime and the marginal tax rates in the country. Some researchers have opposite opinions on this fiscal effect (Kaplan 1989b; Lowenstein 1985). Indeed a going private transaction arouses a large amount of debt used to finance the transaction and creates a considerable additional tax shield.

H1: Leverage and taxation have a positive impact on going private performance.

3.2.2 Free Cash-Flow

Authors argue that for pre-PTP, agency costs are incurred because free cash-flows are spent on projects that do not generate the required positive net present value, Jensen (1986b). These firms will exhibit low growth opportunities and large free cash-flows. The free cash-flows are used to achieve managerial objectives such as increased size and greater peer group standing rather than shareholder wealth maximization. The ability to do this implies ineffective internal corporate governance mechanisms and management would only consider a move away from this situation if faced with an increased threat of hostile take-over. There is an evidence that in the UK hostile takeovers result in a significant increase in the turnover of senior management post-acquisition, (Kennedy and Limmack 1996; Franks and Mayer 1996; Dayha and Powell 1999). It is therefore in the interests of the incumbent management to take a company private and experience increased monitoring rather than risk losing their jobs. Job loss after a hostile take-over would damage their reputation and reduce their value on the executive labor market.

US studies on free cash flow influence in the decision to go private have produced mixed results. Lehn and Poulsen (1989) and Singh (1990) lend support to the free cash-flow hypothesis by reporting that firms going private have greater free cash-flows than firms remaining public. In addition, they found that PTPs exhibited lower sales growth, indicating poorer growth prospects, further supporting Jensen (1986b). However, Kieschnick (1998) reworked Lehn & Poulsen's sample using a weighted logistic regression and found free cash-flows and sales growth to be insignificant. In addition, Opler and Titman (1993) also found no evidence that, individually, either free cash-flows or Tobin's Q influence the decision to go private. However, they found that leveraged buyouts are more likely to exhibit the combined characteristics of low Q ratio and high cash-flow than firms remaining public. Further, Halpern et al (2000) also found no evidence to support the free cash-flow hypothesis. Thus, there is limited evidence that US PTPs exhibit excess free cash-flow and poor growth prospects which suggests that going private is not being driven by the need to return free cash to the shareholders.

H2: Firms that are going private show higher Free Cash Flows than firms remaining public.

3.2.3 Ownership Structure

One aspect of the agency problem that has received little attention is the link between board composition, ownership structures and the PTP decision. In terms of ownership, a US study by Maupin et al. (1984) found that the concentration of ownership among managers and directors was significantly higher in PTPs relative to firms that remain listed. Moreover, monitoring is more difficult with large boards, and buyouts with large syndicates exit sooner as a result, Wright et al. (1995). Indeed, Private Equity firms with significant concentrated ownership have got the incentive and mechanisms to monitor managers through board membership and detailed reporting requirements that go beyond those available to institutional investors in publicly listed corporations, Cumming et al.(2007). In relation to the internal corporate governance mechanisms of listed companies, there has been an increasing international awareness of their role and importance. In the US, the most recent is the Sarbanes Oxley Act (2002). In the UK, number of reports specifically addressed the issue, Cadbury (1992); Greenbury (1995), and proposed that publicly listed companies should adopt a Code of Best Practice, a proposal supported by the London Stock Exchange. Since June 1993, there has been a requirement that listed companies include in their annual reports a statement explaining the extent to which they have adopted the internal governance mechanisms recommended in the Code. The above discussion allows us to propose a number of hypotheses, based on the agency model, to explain the likelihood of a firm going private.

H3: A significant concentrated ownership (Achleitner et al. 2013; Croci and Del Giudice 2014) implies a better governance and a better performance when a firm go private.

3.2.4 Macroeconomic Factors

In addition to classical drivers (firm's characteristics, financial ratios, ownership structure ...), macroeconomic and industry factors may also have an important impact on firm-level returns. However, there are little theoretical references. Indeed companies are exposed to a certain amount of unavoidable economic risks because financial performance is dependent on economic conditions. The common measure of general economic activity is GDP growth which should be positively correlated with buyout returns. As explained by Koller et al. (2005), a company's valuation is directly affected by expectations of its future economic performance.

Some recent studies suggest that industry measures of growth and returns more accurately reflect the fundamentals driving buyout returns, Guo et al. (2011). In addition, the growth rates of individual industries are monitored much less than GDP, and, as a result, industry growth forecasts are likely to be less efficiently priced in transactions than GDP growth forecasts. So, industry growth rates should have a positive impact on buyout returns in addition to the impact of GDP growth rates.

From the study of CMBOR (Center Management for Buy Out Research), the development of LBOs follows the economical cycle. They undergo different financial crisis (1990, 2001–2003) and especially the last one (2007–2008) where the number of LBOs knew a real drop (from 1200 transactions in Europe in 2008 against 400 in 2009). Consequently economic and financial conditions are important for the development of going private transactions. This impact would be important in Asia for different reasons: the economic growth is important (the annual growth amounted to 8.4% in 2000 and reached 14.2% in 2007) and the going private market is still not mature.

H4: Industry growth rates should have a positive impact on buyout returns.

3.3 Institutional Background of Emerging Economies: The Case of Asia

As we study in this article private equity in emerging economies, we have to define "emerging economies". Indeed, the meaning of these words is not the same for researchers belonging to different fields. From Arnold and Quelch (1998), an emerging economy can be defined as a country that satisfies two criteria: a rapid pace of economic development and government policies favoring economic liberalization and adoption of a free-market system. The International Finance Corporation identifies 51 rapid-growth developing countries in Asia, Latin America, Africa and the Middle East as emerging countries. In this article, Asia is selected and particularly three main regions: Central Asia, Far East Asia, South East Asia where 24 countries and five main business sectors are represented (see Tables 3.1 and 3.2).

Table 3.1 Repartition of Asian sample

Areas	Countries	Number
Central Asia	Afghanistan	0
	Armenia	0
	Azerbaijan	0
	Georgia	0
	Iran	2
	Kazakhstan	0
	Kyrgyzstan	0
	Tajikistan	0
	Turkmenistan	0
	Uzbekistan	0
Far East	China	52
	Korea, North	4
	Macau	5
	Mongolia	2
	Taiwan	6
South East	Brunei	1
	Cambodia	2
	East Timor	0
	Indonesia	20
	Laos	0
	Malaysia	27
	Philippines	11
	Thailand	17
	Vietnam	7

Table 3.2 Repartition of business sectors

Business sectors	Number
Construction	11
Finance/Insurance	32
Manufacturing	35
Retail Trade	25
Services	53

3.3.1 Academic Background

Among Asian countries, the first concerned markets were China, India, Ippolito (2007), as well as Japan, Kato and Schallheim (1993). For these three markets, private equity has continued to experience strong capital inflows, Cumming et al. (2010). As for China, funding for private equity deals is mainly provided through offshore holding companies. But, with the development of the local debt market, private equity is gradually moving from growth equity to LBO, as documented by Xiao (2013).

As for India, which is in the midst of structural changes, there are opportunities for a more traditional equity growth. The emphasis on growth leads to many positive changes in management and financial structures, Fang and Leeds (2008). At the industry level, Indian firms have a leading position in services and information technology and a better profitability as compared to Chinese ones, Ippolito (2007). While opportunities are attractive, the risks are high. Hence, according to Fang and Leeds (2008), the retail and the telecom sectors in India were recently beset by unexpected events that undermined the global performance.

As for Japan, the banking system plays an important role in reducing the costs of financial distress, especially when companies maintain longstanding relationships with their bank, Hoshi et al. (1990). However, Kato and Schallheim (1993) documented a positive market reaction after the announcement of private equity introduction. Since then, LBOs were accepted as a form of ownership, encouraged by many elements such as shareholder demand for higher equity returns and divestment from inefficient subsidiaries, Cumming et al. (2010).

3.3.2 Institutional Background

One of the main characteristic of Chinese Governance mechanism is the dominance of state ownership and control (Kato and Long (2006a), (2006b) and (2006c); Chen and AiNajjar 2012). State ownership plays a significant role in bank's management and influences the appointment of directors and the senior management team in the supervisory board in particular. From Firth et al. (2009), there is a negative relationship between government interference in appointing directors and financial performance in China. Pessarossi and Weill (2013) argue that government interference may limit the effectiveness of governance mechanism as this may lead to appointing less profiled (experienced), but loyal, directors in state-owned companies. Lin et al. (2009) find that state ownership may lead to agency problems and has a negative influence on the monitoring role and operating efficiency. Chen and Al-Najjar (2012) find that the higher the level of state ownership, the lower is the supervisory board size and independence.

Ownership structure is one of the main determinants of agency problems. It varies according to the discrepancies in the economic and development stage of each country. The principal-agent problem is very pronounced in the Chinese financial sectors due to government ownership and to the political appointment of directors. In such an environment the primary objective deviates from wealth maximization to social welfare maximization. This may result in corruption and misallocation of resources (Banerjee 1997).

Pessarossi, Weill (2013) find evidence in favor of the influence of central government ownership on the financing choices of firms because central state owned firms are more likely to issue bonds than others and to borrow uniquely on the bond market. Consequently, LBO is not a good solution for the government.

From Li K., Yue H. & Zhao L. (2007), state ownership is positively associated with short-term debt decisions for large firms whereas foreign ownership is strongly and negatively associated with small firms'use of short-term debt. Indeed they show that the negative effect of institutional development on firms' access to long-term debt is mitigated when the level of state or foreign ownership is high.

The attractiveness of Private Equity in China is justified by key driving forces as Groh (2013): first it is the economic activity that explains the importance of Private Equity. However it is not the only reason. Groh (2013) shows a detailed analysis to find the main determinants of Private Equity Attractiveness in Asia. Apart from the economy, the level of unemployment, the level of entrepreneurial tax incentives and administration burden, the security of property rights are the main factors that justify the increase of Private Equity in Asia.

H5: The presence of public authorities among the shareholders has a negative impact on performance.

As Asia presents a specificity in the structure of ownership (a large presence of government ownership), we introduce two variables which represent central and local states. These two variables are coming from the China Security Index. We try to understand if central and local states have the same ownership power.

3.4 Data Sources and Descriptive Statistics

We use several databases (Capital IQ, Worldbank, Center of China Security Index) to analyze the drivers of LBO operating performance in Asia and also activity reports from each firm in order to complete the missing data. We retrieved all the deals from Capital IQ and we selected all the LBO operations with a closed transaction status in Asia from January 2000 to December 2010. The sample period ends in December 2010 in order to assess the performance of delisted firms as private companies in the first 3 years after the going private transactions.

3.4.1 Sample Description

We started with a large sample of 255 Asian companies (Central Asia, Far East, South East) listed on Capital IQ. We added a few criteria to improve our LBO sample analysis.

First, as we are interested in the examination of the post-acquisition performance, it is required that the delisted companies continue operating after the stand alone deal. Consequently, we removed from our sample all takeover targets immediately integrated in the acquirer's legal structure. As we are interested in the observation of companies before and after the delisting decision, takeover targets merged with the bidder do not allow this kind of analysis. After the screening process, 173 companies remained in the sample.

Second, we collected information about the going private deals for all these companies from Capital IQ. Unfortunately, we were not able to find all the needed data for the 173 companies but only for 156 of them. Therefore, the final sample was made of 156 transactions which occurred between 2000 and 2010.

Some regions and countries are more represented than others and provided more data.

In Table 3.2, the most representative business sectors are "services" (34% of the full sample), manufacturing and Finance/Insurance.

3.4.2 Benchmark Comparison

Moreover, as we analyze the impact of LBO transactions, we decided to compare the targets of such transactions to similar companies that did not go through an LBO. We based our peer selection on Capital IQ of listed companies and applied the following matching algorithm to each private observation (similar to Weir et al. 2005; North 2001; Klein and Zur 2009). A matching company i.e. a control firm meets the two following criteria: first we select all public companies which are headquartered in the same country as the going private firms, second we refine our selection by industry. In a first step, we pick all companies which operate in the same two-digit SIC industry. In case there are fewer than five potential matching firms, we enlarge the industry criterion to the one digit SIC code. And in a second step, in order to identify the final matching firm, we employ a size criterion. In particular, we collect the amount of sales of all remaining firms in the fiscal year preceding the going private announcement and the number of employees in full time equivalent in the year prior the LBO transaction. Both criteria (total assets and employees) have to be within the 70–130% range of total assets and number of employees of the corresponding buyout, Barber and Lyon (1996). The firm with the smallest absolute sales deviation from the going private firm is chosen as the matching firm. As a final sanity check, we verify by an examination of the stock prices that our matching firm has stayed public for at least 2 years after the going private announcement. We obtain as the LBO sample 156 firms for the control sample.

3.4.3 Descriptive Statistics

In Table 3.3, we present the definition of variables.

In Table 3.4 we introduce descriptive statistics about the ownership structure, the stock price and the ownership data of our sample of firms delisted following LBO transactions. We can notice that ownership data are collected at the end of the year preceding the delisting announcement. We observe that the level of debt for going private is more important than for non-LBO transactions. It can be explained by the fact that LBO transactions use a significant amount of capital. Therefore the level of leverage is also more important for LBO than for non-LBOs. The different cash-flows generated by LBO operation create a higher level of free cash-flow for LBO than for non-LBOs even if we find a significant level for non-LBOs. We identify that a concentrated shareholder is very significant in LBO sample.

While panel A presents statistics for the full sample, Panel B compares firms delisting following a LBO to firms delisted without LBOs. The acquisition technique is a LBO if Capital IQ considers the deal a leveraged buyout.

3.4.4 Analysis and Discussion

Table 3.4 presents means and medians for the full sample. We observe that financial variables significantly change between year −1 and year +1 and between year −1 and year +3, suggesting a break between the past and the future. The medians results show that delisted firms become smaller, probably due to the asset stripping. Tangible assets decrease significantly, as expected. In fact, delisting often implies downsizing processes to improve efficiency, Shleifer and Summers (1988), Weston et al. (1998).

Some remarks must be mentioned for all variables. First, we observe a decrease in total assets between year −1 and year +3 (−17%). This decrease can be justified by the fact that managers need to reduce their debt level in order to borrow the important amount of debt required for the LBO transaction. Of course, leverage increases strongly. It is the direct effect of LBO transaction. We observe also an Asian specificity in the level of taxation which decreases when a firm makes a LBO operation. The level of free cash flows is constant overtime. However, we notice an important reduction in the ROA level which can be explained by the high level of debt of the LBO firm. However, we notice a constant amount of senior debt. The global cash level decreases. The level of interest rate also decreases and facilitates LBO transactions. We see that the presence of government in the ownership is important, especially for the "large" Far East region. We find that central government has a more important impact in the Far East region than in South East Asia where local governments are determinant.

The performance measurement after the delisting relies exclusively on financial statement data, because stock prices are no longer available once the firm is delisted. Due to the difficulties to obtain reliable financial data, the sample size decreases

Table 3.3 Definitions of variables

	Variables	Definitions
Financial variables	Total assets	A proxy for the size of the company
	Leverage	Ratio between total debts and total assets
	Taxation	All taxes paid by the company during the accounting period scaled by the previous year's total assets
	Free Cash Flow	The sum of the firm's net income plus depreciation scaled by the previous year's total assets
	ROA	The firm's return on assets computed as EBIT (EBITDA) over the firm's total assets at the end of the previous year
Capital structure variables	Divisional	A dummy variable with a value of 1 if the buyout is a division of a larger company and 0 if the buyout comprised a whole company
	Senior debt	The ratio of the amount of senior debt divided by the amount of total debt
	Shareholders	A dummy variable which takes a value of 1 if the shareholder in the firm is large (> 10% of the firm's voting rights) and 0 if the shareholder is dispersed
	Legal status	A dummy variable which takes a value of 1 if the firm is part of a society and 0 otherwise.
Macroeconomic and industry level variables	GDP growth	Annual percentage growth rate of GDP at market prices based on constant local currency. It is calculated from the entry year to exit year
	Industry growth	It is calculated from the entry year to the exit year
	Interest rate	It is an annual rate
	FDI	It refers to direct investment equity flows in the reporting economy. It is the sum of equity capital, reinvestment of earnings, and other capital.
	Market return	The annual return from the beginning of the entry month ont he end of the exit month of share index
Employment	Employees	The number of full time employees of the company
	Profit per employee	The ratio between the firm's profits before taxes divided by the number of employees
Geographical areas	Central Asia	A dummy variable which takes a value of 1 if the firm is located in Central Asia and 0 otherwise
	Far East	A dummy variable which takes a value of 1 if the firm is located in Far East and 0 otherwise
	South East	A dummy variable which takes a value of 1 if the firm is located in South East and O otherwise
Governance	Central State owned	A dummy variable which takes a value of 1 if the borrower is controlled by the central governance and 0 otherwise
	Local State owned	A dummy variable which takes a value of 1 if the borrower is controlled by the local governance and 0 otherwise

Table 3.4 Descriptive statistics

Panel A: Full Sample (312 observations)						
	Year −1		Year 1		Year 3	
	Mean	Median	Mean	Median	Mean	Median
Financial variables						
Total Assets	63456.3	1986.5	57892.4	1832.8**	52411.3	1109.7
Leverage	0.32	0.13	0.41	0.16***	0.52	0.23***
Taxation	2.73	1.64	1.9	0.9***	1.1	0.5***
Free Cash-flows	0.0345	0.0858	0.0395	0.0813***	0.0321	0.0771***
ROA (EBIT)	0.04	0.021	0.05	0.013***	0.02	0.008***
Capital Structure variables						
Divisional	62%***		73%***		69%***	
Senior Debt	0.4367	0.2014	0.4587	0.1987***	0.4432	0.1765***
Shareholders	8.1%	6.9%	7.6%	5.7%***	7.4%	6.3%***
Legal status	6.9%	5.7%	7.2%	6.1%	7.3%	6.0%
Central State Owned	52%					
Local State Owned	38%					
Macroeconomic variables						
GDP growth	10.5%	8.2%	10.3%	7.9%**	10.4%	8.0%**
Industry growth	5.6%	4.3%	4.9%	3.7%**	5.2%	3.9%**
Interest Rate	2.3%*		1.5%*		1.4%*	
FDI	37567.2	2134.5*	32658.2	1935.7**	34679.1	2009.3**
Market return	6.4%	8.6%	8.4%	9.6%**	8.7%	9.8%**

Table 3.4 (continued)

Panel B: LBO vs Non-LBO	LBO						NON LBO					
	Year −1		Year 1		Year 3		Year −1		Year 1		Year 3	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Financial variables												
Total Assets	45673.2	4385.2	38976.3	394.2**	40942.7	4123.6**	75596.5	2250.6	79526.1	2102.4**	71527.8	1331.9**
Leverage	0.32	0.17	0.43	0.20***	0.51	0.28***	0.19	0.07	0.28	0.09***	0.35	0.12**
Taxation	1.7	1.43	1.1	0.97***	0.7	0.85	1.14	1.09***	1.18	1.12***	1.19	1.13***
Free Cash-flows	0.0312	0.0934	0.0298	0.0836***	0.0275	0.0704**	0.0254	0.0791	0.0264	0.0772***	0.0247	0.0757**
ROA (EBIT)	0.014	0.096	0.012	0.078**	0.090	0.065**	0.030	0.031	0.090	0.001***	0.08	0.001**
Capital Structure variables												
Divisional	45%**		52%**		54%**		48%*		59%*		56%	
Senior Debt	0.3290	0.3567	0.4290	0.3214***	0.4798	0.3789***	0.3856	0.1850	0.4259	0.2589**	0.5345	0.3150**
Shareholders	6.2%	5.1%	6.6%	4.8%	6.7%	4.9%	8.5%	7.8%	7.8%	6.5%***	5.8%	5.2%
Legal status	15%		18%		19%		20%		16%		17%	
Central State Owned	62%						35%					
Local State Owned	54%						28%					
Macroeconomic variables												
GDP growth	11.4%	8.2%	9.3%	7.6%	9.1%	7.2%	7.5%	7.4%	7.5%	7.6%**	7.4%	7.3%**
Industry growth	3.4%	2.5%	4.1%	3.6%**	4.5%	3.9%**	2.9%	2.7%	3.4%	2.9%**	3.9%	3.1%**
Interest Rate	3.1%*		2.5%*		1.7%		2.2%*		2.3%*		2.0%*	
FDI	36793.1	1563.6*	35219.4	1245.7*	31905.7	1179.3*	30546.8	1245.9**	29321.4	1158.9*	31275.3	1056.5*
Market return	6.2%	8.3%	7.3%	9.7%***	9.2%	10.4%**	5.9%	7.9%	6.5%	9.1%	6.8%	9.5%

The table reports mean and median of financial variables, capital structure variables and macroeconomic variables for the sample of LBO and for the control sample (non-LBO). The symbols ***, **, * denote statistical significance at the 1%, 5% and 10% level, respectively

Table 3.5 Hypothesis and variables

Hypothesis	Variables
Tax benefit	Leverage
	Taxation
Free Cash Flow	Free cash flow
Ownership	Divisional
	Senior debt
	Shareholders
	Legal status
Macro-economic	GDP growth
	Industry growth
	Interest rate
	FDI
	Market return
Governance	Central State Owned
	Local State Owned

substantially. As a proxy for the modifications in the firm operating performance, we compute the difference between the ROA in the first (third) year after the delisting and the firm's ROA for the last full year in which the company was publicly listed (year −1). In Table 3.6 and more, the dependent variable is the ΔROA. ROA is computed as EBIT over total assets at the beginning of the year. To put it another way, ΔROA is computed as: ROA $_{t+1(3)}$ – ROA $_{t-1}$.

In addition to our key variables, we check the relationship with the following financial variables measured at the end of the year before the delisting: tangible assets, ln(total assets) and ROA. We include the pre-delisting operating performance to measure the persistence of operating performance. We also control for the size effect with the variable Ln (total assets), i.e. the natural logarithm of total assets at the end of the year before the delisting.

To study the performance after the delisting, we use OLS regressions as an econometrical method where Y, the dependent variable is ΔROA and where X_i, the independent variables are Ln (total assets), fixed assets, tangibles, LBO, leverage, taxation, free cash flow, ROA, divisional, senior debt, shareholders, GDP growth, industry growth, interest rate, FDI, market return, central state owned and local state owned.

We introduce also three dummy variables which represent the three geographical areas of our sample.

Table 3.5 summarizes hypotheses and the choice of variables.

Tables 3.6 and 3.7 present three models: model I with all variables, model II checks if we separate ln(total assets) and fixed assets, we get an impact on the quality of the results obtained with the model and model III checks whether ROA has an impact on the quality of the results obtained with the model. The unit used is the million of dollars.

Table 3.6 Change in operating performance between year −1 and year +1

	I	II	III
Ln (Total Assets)	0.032		0.087
	(1.702)*		(1.722)*
LBO	0.022	0.014	0.010
	(3.142)**	(3.456)**	(3.521)**
Leverage	0.023	0.019	0.015
	(4.654)***	(3.009)***	(2.564)***
Taxation	−0.321	−0.490	−0.384
	(3.654)***	(3.109)***	(3.245)***
Free Cash-flows	0.031	0.042	0.049
	(2.784)**	(2.926)***	(2.912)***
ROA (EBIT)	0.314	0.348	
	(2.674)**	(2.743)**	
Divisional	0.069	0.047	0.049
	(2.768)**	(2.905)***	(2.876)**
Senior Debt	0.261	0.298	0.306
	(2.657)***	(2.731)***	(2.769)***
Shareholders	0.763	0.865	0.821
	(2.113)*	(2.219)*	(2.198)*
Legal Status	0.095	0.087	0.092
	(1.682)*	(1.785)*	(1.812)*
GDP growth	0.405	0.532	0.606
	(2.764)**	(2.805)**	(2.913)**
Industry growth	0.178	0.245	0.369
	(2.301)**	(2.787)**	(2.654)**
Interest Rate	0.507	0.514	0.546
	(2.478)**	(2.512)***	(2.874)***
FDI	0.154	0.276	0.198
	(2.765)**	(2.742)**	(2.803)**
Market return	0.389	0.401	0.408
	(2.331)*	(2.489)**	(2.594)**
Central Asia	0.875	0.903	0.894
	(1.004)	(1.012)	(1.045)
Far East	0.453	0.448	0.418
	(2.792)***	(2.714)***	(2.801)***
South East	0.621	0.615	0.613
	(2.422)**	(2.384)**	(2.435)**
Adjusted R^2	0.513	0.504	0.527
Observations	312	312	312

The table reports estimates of OLS regressions where the dependent variable is $\Delta ROA_{(-1,1)}$.$\Delta ROA_{(-1,1)}$ is computed as: ROA $_{t+1}$ – ROA $_{t-1}$. ROA is computed as EBIT over total assets at the beginning of the year. Three models are designed: model I with all variables, model II checks if total assets has an impact on the quality of the results and model III checks if ROA has an impact on the quality of the results. The symbols ***, **, * denote statistical significance at the 1%, 5% and 10% level, respectively

Table 3.7 Change in operating performance between year −1 and year +3

	I	II	III
Ln(Total Assets)	0.045		0.047
	(1.546)*		(1.574)*
LBO	0.162	0.157	0.145
	(2.367)**	(2.427)**	(2.399)**
Leverage	0.035	0.043	0.041
	(3.342)***	(3.297)***	(3.316)***
Taxation	−0.473	−0.491	−0.502
	(2.967)***	(3.122)***	(3.455)***
Free Cash-flows	0.124	0.135	0.127
	(3.115)***	(3.203)***	(3.306)***
ROA (EBIT)	0.291	0.324	
	(2.866)***	(2.983)***	
Divisional	0.065	0.112	0.063
	(2.704)**	(2.589)**	(2.698)**
Senior Debt	0.157	0.215	0.198
	(2.583)**	(2.579)**	(2.642)**
Shareholders	0.094	0.106	0.121
	(2.512)***	(2.486)***	(2.517)***
Legal status	0.189	0.213	0.326
	(2.315)**	(2.416)**	(2.425)**
GDP growth	0.568	0.651	0.709
	(3.189)***	(3.362)***	(3.422)***
Industry growth	0.185	0.206	0.255
	(2.579)**	(2.675)**	(2.609)**
Interest Rate	0.264	0.258	0.241
	(2.345)**	(2.516)**	(2.611)**
FDI	0.104	0.167	0.177
	(2.472)**	(2.389)**	(2.518)**
Market return	0.201	0.279	0.295
	(3.567)***	(3.876)***	(3.906)***
Central Asia	0.773	0.802	0.813
	(1.015)	(1.019)	(1.003)
Far East	0.369	0.379	0.353
	(2.753)***	(2.705)***	(2.792)***
South East	0.548	0.551	0.542
	(2.329)**	(2.371)**	(2.458)**
Adjusted R^2	0.516	0.494	0.535
Observations	312	312	312

The table reports estimates of OLS regressions where the dependent variable is $\Delta ROA_{(-1,3)}$.$\Delta ROA_{(-1,3)}$ is computed as: ROA $_{t+3}$ – ROA $_{t-1}$. ROA is computed as EBIT over total assets at the beginning of the year. Three models are designed: model I with all variables, model II checks if total assets has an impact on the quality of the results and model III checks if ROA has an impact on the quality of the results. The symbols ***, **, * denote statistical significance at the 1%, 5% and 10% level, respectively

3.5 Results

We based our measure of performance after the delisting on the difference between the ROA in years 1 and 3 after the delisting. Therefore, the dependent variable is the variation of ROA (ΔROA): ROA $_{t+1}$ – ROA $_{t-1}$ when we investigate short-term effects and ROA $_{t+3}$ – ROA $_{t-1}$ when we look at long-term effects.

ROA is defined as EBITDA over total assets. We study the relation betweenΔROA and independent variables before the delisting in a first time and 3 years after the delisting in a second time. Candidate independent variables are: financial variables, (ln (total assets) which is the natural logarithm of total assets at the end of the year before the delisting, a proxy for the size of the target; fixed assets; tangibles; leverage; taxation; free cash-flows; ROA; and cash reserves), capital structure (divisional, senior debt, shareholders), macroeconomic variables (GDP growth, industry growth, interest rate, foreign direct investment, market return), government presence (central and local stakes owned) and the level of employment (employees, profit per employee).

Two models are created:

The first model deals with the main drivers of LBO performance. We divided this model into two parts: first the main drivers before the going private transaction and immediately after. The second part studies the period before the transaction and 3 years after.

$$\begin{aligned}
\Delta ROA\ \ (-1,1) = {} & \beta_1 \ln(\textit{total assets}) + \ \ \beta_2\ \textit{LBO} + \beta_3\ \textit{Leverage} + \beta_4\ \textit{Taxation} \\
& + \beta_5\ \textit{FreeCashFlow} + \beta_6\ \textit{ROA} + \beta_7\ \textit{Divisional} + \beta_8\ \textit{SeniorDebt} \\
& + \beta_9\ \textit{Shareholders} + \beta_{10}\ \textit{Legal Status} + \beta_{11} \textit{GDP growth} \\
& + \beta_{12}\ \textit{Industry growth} + \beta_{13}\ \textit{Interest rate} + \beta_{14}\ \textit{FDI} \\
& + \beta_{15}\ \textit{Market return} + \beta_{16} \textit{Central Asia} + \beta_{17} \textit{Far East} \\
& + \beta_{18} \textit{South East} + \varepsilon_i
\end{aligned}$$

$$\begin{aligned}
\Delta ROA\ \ (-1,3) = {} & \beta_1 \ln(\textit{total assets}) + \ \ \beta_2\ \textit{LBO} + \beta_3\ \textit{Leverage} + \beta_4\ \textit{Taxation} \\
& + \beta_5\ \textit{FreeCashFlow} + \beta_6\ \textit{ROA} + \beta_7\ \textit{Divisional} \\
& + \beta_8\ \textit{SeniorDebt} + \beta_9\ \textit{Shareholders} + \beta_{10}\ \textit{Legal Status} \\
& + \beta_{11} \textit{GDP growth} + \beta_{12}\ \textit{Industry growth} + \beta_{13}\ \textit{Interest rate} \\
& + \beta_{14}\ \textit{FDI} + \beta_{15}\ \textit{Market return} + \beta_{16} \textit{Central Asia} + \beta_{17} \textit{Far East} \\
& + \beta_{18} \textit{South East} + \varepsilon_i
\end{aligned}$$

We are interested in the effect of ownership on the performance of going private transactions. We study the impact of central and local governments on each geographical area. We use all periods to get a better impact, but we don't split the analysis between short term (−1,+1) and long term (−1,+3).

$$\Delta ROA\ (-1,3) = \beta_1 \mathrm{CA}^{*}\mathrm{CSO} + \beta_2\, CA^{*}LSO + \beta_3\, FE^{*}CSO + \beta_4\, FE^{*}LSO + \beta_5\, SE^{*}CSO + \beta_6\, SE^{*}LSO + \varepsilon_i$$

where CA is Central Asia, FE is Far East, SE is South East; CSO is Central State Owned, LSO is Local State Owned.

3.5.1 OLS Model

Table 3.6 shows the results issued from the short-term forms of OLS model estimations. We observe that the history of the firm variable has a positive and significant effect on performance of going private transactions. Indeed, in Asia, we have a large number of small firms and family entrepreneurship is dominant (Sannajust 2009). We demonstrate that when a family shareholder initiates a going private transaction, this affects positively the firm's operating performance. As said by Olivier Carcy, the Geneva based Global Head of Private Equity at Crédit Agricole's private banking unit: "Some people say Asia isn't primed for leveraged buyouts because of the prevalence of family-run companies. But I think it's just a matter of maturity. Once the financial markets develop to support leveraged buyouts then they'll naturally emerge".

Indeed this result could be interpreted as a signal of asymmetric information. A large shareholder (in this case it is the family shareholder) takes a firm private because it has better quality information about firm's profitability. We can explain this finding by the agency theory: the reduction of agency conflicts between small and large shareholders generates an improvement in the firm's performance. After the delisting, family shareholders have additional incentives to run the firm efficiently because they often invest their own financial resources to buyout minorities and get the control of the LBO transaction, since these acquisitions are rarely financed by a debt increase. Therefore, we can deduce that the level of performance depends on the owner's post-delisting situation.

3.5.2 Introduction of LBO Dummy Variable

LBO dummy shows a positive and statistically significant coefficient. Indeed LBO as a technique of acquisition influences positively the results of performance. Leverage presents also significant and positive results. Greater availability of debt and lower interest rates on borrowing are associated with higher leverage in buyout financing structure, Axelson et al. (2012). Leverage should lead to increased firms-level holding periods and equity returns particularly in successful buyouts, because of pressure to meet service debt requirements. Free Cash-flow has a positive and significant effect on performance for year 1 and especially for year 3. Indeed, LBOs have higher levels of Free Cash-flows, in average than non-LBOs. The excess cash owned by LBO can repay the debt. The positive relation is confirmed by other studies Wright et al. (2006) but their results are not statistically significant. Becker and Pollet (2008) for US sample find a positive and significant link between going private and the level of free cash-flows. For taxation, we remark a higher level of tax for LBOs than for non-LBOs. In general the result for taxation is not significant, Wright et al.(2006), for Europe and the USA samples. This could be explained by the fact it is a new trend and we also notice high growth rate for LBOs and large flows of private equity. We can assume that the post-LBO growth can be explained by an expansion on international markets.

As far as macroeconomic and industry variables are concerned, our findings indicate that industry growth has a significantly positive impact on performance which is the same results as Guo et al. (2011). As expected, GDP growth is significant. Market return, which is measured by the market adjusted stock price performance in the calendar year before the announcement, presents a positive and significant result before the going private transaction and indicates that the stock market was able to forecast future firm's performance. Asian markets confirm their infatuation for LBO transactions. In contrast with the USA, LBOs are new in Asia and the market reaction is different.

In Table 3.7, we focus on the relationship between going private and long-run performance. We use ΔROA $_{(-1,3)}$ as a dependent variable. We confirm the results obtained in Table 3.6; the increase in operating performance when a large shareholder takes over the firm is permanent.

We also demonstrate that adjusted R^2 which represents the quality of the model are lower for years -1, $+1$ than for years -1, $+3$ (51.3% against 51.6% for model I, 50.4% against 49.4% for model II and 52.7% against 53.5% for model III). These differences do not interfere with the results of our regressions. In general we show that the impact of operating performance is more important for 1 year before and 1 year after the transaction. It can be explained by the fact that 1 year after the delisting, firms are more flexible: all constraints and costs incurred by the exchange do not apply anymore. Financial results increase. However, as we know, LBO transactions imply the extensive use of debt. Therefore, managers are very careful because the firm has to repay the loan in due time. This is an argument to explain the lower results obtained for adjusted R^2 during the performance years -1 and $+3$. However, in Asia, the impact 3 years after the LBO is also important for the firm. We can explain this result by the fact that LBO is central for a firm and its managers give the same importance to create value at the beginning of the transaction and during the three following years. In the USA and in Europe we don't have the same interpretation, because managers give a priority to short term results i.e. a great importance to the beginning of the period (1 year after) and less after.

3.5.3 Introduction of Geographical Area Dummy Variables

Tables 3.6 and 3.7 contain three geographical dummy variables that represent the three Asian regions of our sample. We have a geographical cluster where Far East and South East Asia are the most representative areas (Table 3.1). If we don't introduce these variables, our result would not represent the real picture.

We find that Central Asia has a positive but not significant impact on the performance. It is not surprising as the number of going private is very low. On the contrary, Far East and South East Asia present positive and significant results. The quality of these results is explained by the large number of transactions which occurred. China, Indonesia, Malaysia, Philippines, Thailand are the five main countries where going private operations took place. We also notice a better level of significance for Far East where China is the most representative country. The level of growth, the attractiveness of the financial markets, the size of the country are such that China has an important attractiveness factor.

Finally, these results should be interpreted taking into account the location of going private firms. Location has a positive and significant impact on going private performance.

3.5.4 Geographical Areas and Governance (Table 3.8)

Firms ownership in Asia is characterized by the large presence of public authorities stakes (Pessarossi and Weill, 2013; Li K., Yue H., Zhao L., 2007). To obtain better results in our paper we decided to use variables from the Center of China Security Index. Central state and local states ownerships variables show if their presence by geographical area is significant or not and if their influence is positive or negative. We find that the large presence of public authorities has a negative and significance influence (Firth et al. 2009; Lin et al. 2009) on the performance of going private firms. These results are justified by the fact this presence of public authorities creates some agency conflicts. The managers of these firms are more motivated to issue debt on the bond market than to borrow to the banks.Moreover the location is also an important factor. In a "large" region as Fast East, central government has a more important impact than in the South East region where the size of regions is smaller.

3.5.5 Efficiency and Profitability Impacts (Table 3.9)

Going private transactions imply also some restructuration to improve the firm's efficiency. Thus, we introduce another variable, the employment. We study the LBO's effect on employees. As we know, a LBO transaction implies restructuration and financial investments to be successful in the delisting process. Therefore,

Table 3.8 Geographical areas and public authorities

	I	II
Central Asia*CSO	−0.352	
	(−1.004)	
Central Asia*LSO		−0.535
		(−1.112)
Far East*CSO	−0.233	
	(−2.851)***	
Far East*LSO		−0.301
		(−2.112)**
South East*CSO	−0.342	
	(−2.215)**	
South East*LSO		−0.243
		(−2.421)***
Adjusted R^2	0.456	0.461
Observations	156	156

The table shows the impact of the presence of public authorities in each geographical areas in our sample. The symbols ***, **, * denote statistical significance at the 1%, 5% and 10% level, respectively

Table 3.9 The level of employment

	Year −1		Year 1		Year 3	
	Mean	Median	Mean	Median	Mean	Median
PANEL A: FULL SAMPLE						
Employees	1245	204**	1105	183**	1389	231***
Profit per employees	198.6	59**	204.6	62***	285.1	86***
PANEL B: LBO vs. NON-LBO						
Employees	2145	457**	1987	214**	2345	632***
Profit per employees	187.2	65**	308.4	96***	510.3	113***

The table reports mean and median of employees, profit per employees before the delisting (year −1), the year after (year 1) and 3 years later (year +3) for the sample. Employees represent the number of full time employees of the company. Profit per employees is the ratio between the firm's profits before taxes divided by the number of employees. The symbols ***, **, * denote statistical significance at the 1%, 5% and 10% level, respectively

efficiency is the main goal of LBO transactions. Indeed going private transactions imply the improvement of firm's efficiency by restructuring the firm after the LBO, Shleifer and Summers (1988), Weston et al. Kaplan (1989a). Consequently, the efficiency improvements are obtained through cost cutting in assets and employment, Kaplan (1989a), Smith (1990), Harris et al. (2005). To test this idea, we use two variables: employee which is the number of total employees of the company and the profit per employee ratio which is calculated with the firm's profit before taxes divided by the number of employees. With these two variables we can analyze the effect of this restructuring process on the firm's workforce and its efficiency.

Panel A shows that there is a decrease in the number of employees after the first year of delisting whereas the profit per employee increases. We can suggest that a reduction of employment leads to an improvement in productivity and later as the firm after delisting wants to reduce the incidence of the cost of employees to a workforce reduction and/or to a decrease in the wage per hour, Kaplan (1989a), Smith (1990), Harris et al. (2005). We conclude that firms use going private transactions to restructure their workforce through the number of employees and their cost.

We also notice two opposite results: we get a significant increase in profitability per employee just after the going private whereas we find a decrease in employment level. This result is similar to other studies about LBOs and efficiency in Europe, Boucly et al. (2011), Harris et al. (2005), Cumming et al. (2007). We conclude that a going private transaction as an acquisition technique allow firms to restructure their workforce. This has a positive impact on the firm's productivity with an increase in the profit per employee. As Shleifer and Summers (1988) explained, it is easier to break implicit contracts with employees for a new owner.

However, we notice another relevant result, an increase in the number of employees 3 years after the delisting contrary to other countries where this number decreases each year after the delisting. It is the same interpretation we already mentioned before: managers and shareholders in Asia show a specific behavior, they measure the performance and the quality of the management of the firm on a long time horizon. As we confirm that the profit per employee also increases, it means, in average, firms develop their activity and need more staff to meet a growing demand.

3.6 Conclusion

In this paper, we contribute to private equity research and more precisely to the improvement of knowledge in Asian LBO transactions. The drivers of performance were identified through the analysis of 156 operations. The increase in the level of foreign investments and in the number of equity capital operations and more generally the high growth rate of the economies explain the choice of Asian countries for our research.

While most of the papers available on LBOs explain the operation effect around the delisting date, we studied the impacts before and after the delisting (1 year before and 3 years after). We included macroeconomic variables to take into account GDP growth rates and volatility and also a control sample for non-LBO transactions (Sannajust et al. 2015).

We found that buyouts create value, reduce agency costs, generate a shift from a managerial to an entrepreneurship mindset and lead to an increase in growth for the economy (Chevalier and Sannajust 2011). In the LMBO case, managers resume the company in their own direction and are involved in the decision process and motivated by the issue. The introduction of a "divisional variable" in the model demonstrates that divisional buyouts create more value through acquisitions than integrated company buyouts. Information asymmetries between existing and new management teams explain this difference in performance. Other analysis including leverage, ROA, market return and shareholders characteristics variables confirm the preceding result. We don't validate the non-significant result obtained for the taxation variable by several authors, Wright et al. (2006); in our analysis, taxation has a positive impact on LBO transactions. Indeed, LBO processes imply large financial flows and tax consolidation plays an important role.

In our model, macroeconomic variables show a positive and significant influence on value creation (industry growth and GDP growth for example). We conclude that a positive macroeconomic environment is necessary for the development of LBOs and also of value creation on LBO transactions. Economic and financial academics explain that LBOs are one of the processes used to implement drastic "cost cutting" measures that the target management is reluctant to enforce and act as growth engines. We validate this hypothesis because we observe the number of employees decreases over the years while the net earnings per employee increases; this result means that LBO transactions imply a workforce restructuration. We also find that LBOs have higher financial performance (ROA, level of assets . . .) than the control sample.

The introduction of geographical dummy variables shows that Far East is the main region as far as the number of going private transactions is concerned. This is validated by the econometrical analysis. The South East region also shows a significant result. Only Central Asia doesn't have significant results. Negative and significant results (agency conflicts and asymmetric information are the main reasons) are explained by the presence of central and local public authorities in the equity structure.

When we analyze the relationship between financial performances of LBOs, our study reveals that, unlike in the USA and in Europe where the operating performances are only important 1 year before and 1 year after the transaction, the impact for Asian firms stays at a high level 3 years after the LBO. In Asia, managers give the same importance to value creation any time and demonstrate a constant behaviour different from their US or European counterparts.

To sum up, this paper brings additional evidence in favor of "the LBO better performance argument" in another region of the world (after the USA, Europe and Latin America) and considers new independent variables as drivers of operating performance. Macroeconomic variables show an impact as important as governance factors on LBO value creation. The presence of public authorities as shareholders has a negative impact on going private operations due to the agency costs created. The characteristics of the debts included in the balance sheets (maturity, fixed or variable interest rates for example) are not available in our data basis. A test including this information could bring other elements of explanation. A comparative study between emerging countries including investigations on heterogeneities could also be a topic for future research.

References

A. Achleitner, A. Betzer, M. Goergen, B. Hinterramskogler, Private Eqtuiy acquisitions of continental European firms : the impact of ownership and control on the likelihood of being taken private. European Financial Management **19**(1), 72–107 (2013)

D.J. Arnold, J.A. Quelch, New strategies in emerging economies. Sloan Manag. Rev. **40**(1), 7–20 (1998)

U. Axelson, T. Jenkinson, P. Stömberg, M.S. Weisbach, Borrow cheap, buy high? The determinants of leverage and pricing in buyouts. J. Financ. **68**(6), 2223–2267 (2012)

S.C. Bae, R.J. Hendershott, H. Jo, Choosing an organizational form: leveraged buyouts versus leveraged recapitalizations. Working Paper Bowling Green State University, (2000)

Banerjee, A theory of misgovernance. Q. J. Econ. **112**(4), 1289–1332 (1997)

Barber, Lyon, Detecting abnormal operating performance: the empirical power and specification of test statistics. J. Financ. Econ. **41**, 359–399 (1996)

B. Becker, J. Pollet, The Decision to go private. Working paper, (2008)

A.N. Berger, G.F. Udell, The economics of small business finance: the roles of private equity and debt markets in the financial growth cycle. J. Bank. Financ. **22**, 613–673 (1998)

Q. Boucly, D. Sraer, D. Thesmar, Job creating LBO. J. Financ. Econ. **102**(2011), 432–453 (2011)

Cadbury Committee, *Report of the Committee on the Financial Aspects of Corporate Governance* (Gee, London, 1992)

C.H. Chen, B. Al-Najjar, The determinants of board size and independence: evidence from China. Int. Bus. Rev. **21**(5), 831–846 (2012)

A. Chevalier, A. Sannajust, "Why do firms go private?" Bankers, Markets and Investors. n°113 (2011)

D.G. Croci, Delistings controlling shareholders and firm performance in Europe. Eur. Financ. Manag. **20**(2), 374–405 (2014)

D. Cumming, D.S. Siegel, M. Wright, Private equity, leveraged buyouts and governance. Finance **13**, 439–160 (2007)

D.J. Cumming, U. Walz, Private equity returns and disclosure around the World. J. Int. Bus. Stud. **41**, 727–754 (2010)

D.J. Cumming, G. Fleming, S. Johan, M. Takeuchi, Legal protection, corruption and private equity returns in Asia. J. Bus. Ethics **95**, 173–193 (2010)

J. Dahya, R. Powell, Top management changes following hostile and friendly take-overs ACCA Research Report, 61. (London, 1999)

H. DeAngelo, L. DeAngelo, E. Rice, Going private: minority freeze outs and stockholders' wealth. J. Law Econ. **27**, 367–402 (1984)

C. Diller, C. Kaserer, What drives cash-flow based European private equity returns? – Fund inflows, skilled GPs and/or risk? European Financial Management **15**, 643–675 (2009)

M. Faccio, L.H.P. Lang, The ultimate ownership of western European corporations. J. Financ. Econ. **65**, 365–395 (2002)

E.F. Fama, Agency problems and the theory of the firm. J. Polit. Econ. **88**, 134–145 (1980)

E. Fama, M. Jensen, Separation of ownership and control. J. Law Econ. **40**, 301–349 (1983)

L. Fang, R. Leeds, Indian private equity cases: introduction, in *Globalization of Alternative Investments Working Papers Volume 1: The Global Economic Impact of Private Equity Report 2008*, (World Economic Forum, Geneva, 2008), pp. 141–142

C. Firth, M.A. Charleston, S. Duffy, B. Shapiro, E.C. Holmes, Insights into the evolutionary history of an emerging livestock pathogen: Porcine Circovirus 2. J. Virol. **83**, 12813–12821 (2009)

J. Franks, C. Mayer, Hostile take-overs and the correction of managerial failure. J. Financ. Econ. **40**, 163–181 (1996)

G.M. Frankfurter, E. Gunay, Management buy-outs: the sources and sharing of wealth between insiders and outside shareholders. Q. Rev. Econ. Finance **32**, 82–95 (1992)

R. Greenbury, *Directors' Remuneration: Report of the Study group Chaired by Sir Richard Greenbury* (Gee, London, 1995)

S. Guo, E.S. Hotchkiss, W. Song, Do buyouts (still) create value? J. Financ. **66**, 479–517 (2011)

Groh, Private Equity in Emerging markets, Working papers (2013)

P. Halpern, R. Kieschnick, W. Rotenberg, Why firms engaged in levered recapitalisation rather than levered buyout, Working Paper University of Texas at Dallas (2000).

R. Harris, D. Siegel, M. Wright, Assessing the impact of management buyouts on economic efficiency: plant-level evidence from the United Kingdom. Rev. Econ. Stat. **87**, 148–153 (2005)

T. Hoshi, A.K. Kashyap, D. Sharfstein, The role of banks in reducing the costs of financial distress in Japan. J. Financ. Econ. **27**, 67–88 (1990)

IFC, International Finance Corporation, World Bank Group (2012)

R. Ippolito, Private Equity in China and India. J. Priv. Equity. **10**(4), 36–41 (2007)

M.J. Jensen, W.H. Meckling, Theory of the firm: managerial behavior, agency costs and ownership structure. J. Financ. Econ **13**, 305–360 (1976)

M.C. Jensen, W.H. Meckling, "Divisional performance measurement", Harvard Colloquium on field studies in accounting, June 18–20, in *Foundations of Organizational Strategy,* ed. by M.C. Jensen (Harvard University Press, Cambridge, (1986a) 1998), pp. 345–361

M. Jensen, The take-over controversy: analysis and evidence. Middle. Corp. Financ. J. **4**, 323–329 (1986b)

S.N. Kaplan, The effects of management buyouts on operating performance and value. J. Financ. Econ. **24**, 217–254 (1989a)

S.N. Kaplan, Management buyouts: evidence on taxes as a source of value. J. Financ. **44**, 611–632 (1989b)

S.N. Kaplan, P.J. Strömberg, Leveraged buyouts and private equity. J. Econ. Perspect. **23**(1), 121–146 (2009)

K. Kato, J. Schallheim, Private equity financings in Japan and corporate grouping. Pac. Basin Financ. J. **1**(3), 287–307 (1993)

T.K. Kato, C.X. Long, CEO turnover, firm performance, and enterprise reform in China: evidence from micro data. J. Comp. Econ. **34**(4), 796–817 (2006a)

T.K. Kato, C.X. Long, Executive compensation, firm performance, and corporate governance in China: evidence from firms listed in the Shanghai and Shenzhen Stock Exchanges. Econ. Dev. Cult. Chang. **54**(4), 945–983 (2006b)

T.K. Kato, C.X. Long, Executive turnover and firm performance in China. Am. Econ. Rev. **96**(2), 363–367 (2006c)

V. Kennedy, R. Limmack, Take-over activity, CEO turnover, and the market for corporate control. Journal of Business Finance and Accounting **23**, 267–293 (1996)

R. Kieschnick, Free cash-flow and stockholder gains in going private transaction revisited. Journal of Business Finance and Accounting **25**, 187–202 (1998)

O. Kini, W. Kracaw, S. Mian, Corporate takeovers, firm performance and board composition. Finance **1**, 383–412 (1995)

A. Klein, E. Zur, Entrepreneurial shareholder activism: hedge funds and other private investors. J. Financ. **64**(1), 187–229 (2009)

T. Koller, M. Goedhart, D. Wessels, *Valuation: Measuring and Managing the Value of Companies* (Wiley, Hoboken, 2005)

R. La Porta, F. Lopez-de-Silanes, A. Shleifer, Corporate ownership around the world. J. Financ. **54**, 471–517 (1999)

K. Lehn, A. Poulsen, Free cash-flow and stockholder gains in going private transactions. J. Financ. **XLIV**, 771–787 (1989)

C. Lin, Y. Ma, D. Su, Corporate governance and firm efficiency : evidence from China's publicly listed firms. Manag. Decis. Econ. **30**(3), 193–209 (2009)

A. Ljungqvist, M. Richardson, D. Wolfenzon, The investment behavior of buyout funds: theory and evidence. New York University Working Paper FIN-07-020, (2008)

L. Lowenstein, Management buyouts. Columbia Law Rev. **85**, 730–784 (1985)

A. Malenko, N. Malenko, A theory of LBO activity based on repeated debt-equity conflicts. J. Financ. Econ. **117**, 607–627 (2015)

H.G. Manne, Mergers and the market for corporate control. J. Polit. Econ. **73**, 110–120 (1965)

R.J. Maupin, Bidwell, Ortegren, An empirical investigation of the characteristics of publicly quoted companies which change to closely held ownership through management buy-outs. J. Bus. Financ. Account. **11**, 345–359 (1984)
R. Morck, A. Shleifer, R.W. Vishny, Characteristics of targets of hostile and friendly take-overs corporate take-overs: causes and consequences, in *Corporate Take-overs: Causes and Consequences*, ed. by A.J. Auerbach (University of Chicago Press, Chicago, 1988)
S.C. Myers, Determinants of corporate borrowing. J. Financ. Econ. **5**, 147–175 (1977)
S.A. Naqi, H. Samanthala, Venture capital or private equity? the Asian experience. Bus. Horiz. **50**(4), 335–344 (2007)
E. Nikoskelainen, M. Wright, The impact of corporate governance mechanisms on value increase in leveraged buyouts. Finance **13**, 511–537 (2007)
D.S. North, The role of managerial incentives in corporate acquisitions: evidence from the 1990s. Finance **7**, 125–149 (2001)
A. Oberli, Private equity in emerging markets: drivers in Asia compared with developed countries. J. Priv. Equity. **17**(3), 45–61 (2014)
T. Opler, S. Titman, The determinants of leveraged buyout activity: free cash-flow vs. financial distress costs. J. Financ. **48**(5), 1985–1999 (1993)
J. Paglia, M. Harjoto, The effects of private equity and venture capital on sales and employment growth in small and medium-sized businesses. J. Bank. Financ. **47**, 177–197 (2014)
P. Pessarossi, L. Weill, Do capital requirements affect bank efficiency ? Evidence from China, BOFIT Discussion Paper, 28. Helsinki Bank of Finland, (2013a)
P. Pessarossi, L. Weill, Choice of corporate debt in China : the role of state ownership. China Econ. Rev. **26**, 1–16 (2013b)
L. Phalippou, M. Zollo, What drives private equity fund performance? Wharton Financial Institutions Center Working Paper #05–41. (2005)
A. Sannajust, “Motivations and performance of public to private transactions: an international study”, PhD in Finance, University of Clermont-Ferrand, 310 (2009)
A. Sannajust, M. Arouri, A. Chevalier, Drivers of LBO operating performance: an empirical investigation in Latin America. Eur. Bus. Rev. **27**(2), 102 (2015)
A. Shivdasani, Board composition, ownership structure and hostile takeovers. J. Account. Econ. **16**, 167–198 (1993)
A. Shleifer, C.H. Summers, Breach of trust in hostile takeovers. Chapter 2, in *Corporate Takeovers: Causes and Consequences*, ed. by A.J. Auerbach (University of Chicago Press, Chicago, 1988)
H. Singh, Management buyouts and shareholder value. Strateg. Manag. J. **11**, 111–129 (1990)
A. Smith, Capital ownership structure and performance: the case of management buyouts. J. Financ. Econ. **13**, 143–165 (1990)
S. Toms, M. Wright, Divergence and convergence within Anglo-American Corporate Governance Systems: evidence from the US and UK, 1950–2000. Bus. Hist. **47**, 267–295 (2005)
C. Weir, D. Laing, M. Wright, Undervaluation, private information, agency costs and the decision to go private. Appl. Financ. Econ. **15**, 947–961 (2005)
J.F. Weston, K.S. Chung, J.A. Siu, *Takeovers, Restructuring and Corporate Governance*, 2nd edn. (Prentice-Hall, New York, 1998)
M. Wright, S. Thompson, K. Robbie, P. Wong, Management buyouts in the short and long-term. J. Bus. Financ. Account. **22**, 461–482 (1995)
M. Wright, R.E. Hoskisson, L.W. Busenitz, J. Dial, Entrepreneurial growth through privatization: the upside of management buyouts. Acad. Manag. Rev. **25**, 591–601 (2000)
M. Wright, L. Renneboog, T. Simons, Public-to-Private transactions in the UK, Working paper, Tilburg University, (2006)
L. Xiao, *From Growth Equity to Leveraged Buyout: Making Private Equity Investments in China* (MIT Libraries, 2013), p. 101

Chapter 4
Time Varying Correlation: A Key Indicator in Finance

Rita L. D'Ecclesia and Denis Kondi

Abstract Correlations between different asset returns represent a crucial element in assets allocation decisions and financial engineering. In commodity markets, where prices result non stationary and returns are only mean stationary, a time varying measure of correlation has to be used. According to the prevailing literature, correlations among different markets are higher during recessions than during expansion periods. Portfolio managers to shield investors from stock markets declines used to invest in commodities which historically were considered poorly correlated with stock markets and providing a good hedge in the long run. In the last decade correlations between commodities and stock returns have dramatically changed. The aim of the paper is to address the issue of the correlation measurement in presence of non stationarity and structural breaks in market variables. We compare the Historical Rolling Correlation and the Dynamic Conditional Correlation methods and show how each estimator may provide useful information given a specific structure of the data. Some interesting relationships are also highlighted among markets where no correlations were expected and viceversa. We also show that information provided by the correlation measures can be used to identify structural breaks in the original variables.

Keywords Time varying correlation • Structural breaks • Commodities markets • Energy markets

JEL Classification Numbers: C32-Q40

R.L. D'Ecclesia (✉)
Dipartimento di Scienze Statistiche, University of Rome La Sapienza, Rome, Italy
e-mail: rita.decclesia@uniroma1.it

D. Kondi
KPMG Milan, Italy

G. Consigli et al. (eds.), *Handbook of Recent Advances in Commodity and Financial Modeling*, International Series in Operations Research & Management Science 257,
DOI 10.1007/978-3-319-61320-8_4

4.1 Introduction

Understanding how asset prices behave and interact in financial markets has become a key issue for scholars and practitioners. Given that the economy is an interconnected set of economic agents, sometimes modelled as a continuously evolving general equilibrium system, we may expect that corporate earnings and asset prices be correlated. The correlation structure across assets is a key feature of the portfolio choice problem because it is instrumental in determining risk as well as being a key factor for pricing exotic derivatives as inter-commodity spread options. Estimating the correlation structure of assets traded in global markets and using this to select superior portfolios is one of the main goal of financial investors or hedge fund companies. It is especially difficult when it is recognized that these correlations vary over time. The optimal hedge can be set only if price dynamics are understood and their volatility properly measured. In addition, derivatives such as options are now routinely traded not only on individual securities, but also on baskets and indices. The pricing of these derivative contracts depend on the prices of the component assets and of their correlations. A market for correlation swaps[1] has recently developed that allows traders to take a position to speculate on or hedge risks associated with the observed average correlation of a collection of underlying products. Structured products form a very large class of derivatives that are sensitive to correlations. In this context it is important to identify the adequate methods for estimating correlations when stationarity of asset prices or their returns cannot be assumed. In some cases the *Historical Rolling Correlation* (HRC) method may provide a simple and accurate measure of time varying correlations, in others the *Exponential Smoothing Correlation* (ESC) or the multivariate GARCH and *Dynamic Conditional Correlation* (DCC) methods may result more accurate in presence of complex structures of the data.

Recently, measuring correlations between different kind of asset classes has become a major task for traders and fund managers. Portfolio managers who aim to shield investors from stock markets declines invest in commodities which historically were considered poorly correlated to stock markets. However, recent crisis in the global economy, as the GFC or the downgrading of the US sovereign debt, or the European sovereign crisis, have caused structural changes in the relationship between the various markets. We can observea disparity in the performance of equity and commodity markets as an indirect endorsement of the government's efforts to shore the domestic economy without stoking inflation. As a response to the Central Banks (FED, ECB) bond purchasing programs, fund managers boosted exposure to commodities amid fears that Central Banks interventions would fan inflation by injecting excessive amounts of money into the economy.

[1]A correlation swap is an over-the-counter financial instrument that allows an investor to speculate on the correlation, of underlying securities.

The aims of this paper is twofold: first, to provide an accurate description of the various time varying correlation measures and to identify the most adequate estimator when dealing with assets which present complex dynamics and different distribution features; second, to show how choosing an adequate correlation measure may lead to a correct estimate of the relationship between markets which present changing structure over time, for instance commodity markets and financial markets, or assets which have been experiencing several structural changes, as well as the crude oil and the natural gas markets.

To select the best correlation measure we need to create experimental data and implement a known correlation structure. We compute different correlation estimators: (1) the HRC choosing different sizes of the rolling windows, (2) the ESC and (3) the DCC. Different error metrics, as the Mean Square Error, the Mean Absolute Error, the Mean Absolute Percentage Error and the Normalised Mean Square Error allow to rank the various estimators.

Using the $HRC - 250$ days we estimate the correlations between natural gas and crude oil prices to show how the use of an accurate measure may provide important information on the relationship existing between the two major fossil fuels traditionally expected to be correlated. We also estimate the DCC to measure the existing relationship between the commodity and the financial markets which have shown major changes in the last decade. We find that the HRC provide an accurate correlation measure in the case of data which present structural breaks, while the DCC is the most adequate estimator for heteroskedastic data.

The paper is organized as follows, the second section reports some relevant literature on the topic; Sect. 4.3 describes and compares the different correlation estimators; Sect. 4.4 presents the measured correlation between different markets (commodities versus stocks and bonds), and different assets (crude oil versus natural gas). Section 4.5 provides some concluding remarks.

4.2 Recent Literature

Measuring correlation in finance is a critical task. Correlation between two asset returns is the basis for any portfolio selection or any diversification strategy. It has to capture the dynamic structure of the daily prices or returns. The standard Pearson correlation coefficient is not appropiate due to the presence of non stationary variables or only mean stationary variables. Financial series are almost always non stationary. Their returns[2] are stationary in mean but show the presence of heteroskedasticity. Countless results of researches dealt with the problem of reliable estimates of correlations between financial variables. Simple methods such as the Historical Rolling Correlations and the Exponential Smoothing Correlation are widely used. More complex methods, such as varieties of multivariate generalized

[2]Returns are calculated with the following formula: $r_t = ln(P_t/P_{t-1})$.

autoregressive conditional heteroskedasticity (GARCH) or stochastic volatility, and the Dynamic Conditional Correlation method have been extensively investigated in the econometric literature and are largely used by practitioners. Just to cite some among the most relevant contributions (Bollerslev et al., 1988, 1994; Bollerslev, 1990; Engle and Mezrich, 1996; Engle et al., 1990; Ding and Engle, 2001; Engle, 2002).

Several studies try to assess the correlations among various markets. A first group of researches dealt with measuring the correlation between commodity markets and other financial markets using data referred to periods before the *Global Financial Crisis* of 2008 (Gorton et al., 2007; Erb and Campbell, 2006; Büyükşahin et al., 2009). They find no significant correlation between the two markets, studying the correlation between commodity indexes and equity indexes in the U.S. market for a period of 17 years (from January 1991 through May 2008). The returns on investible commodities showed a different pattern from that detected for equity indexes, and the relationship had not changed significantly during the period 1991–2008. They conclude that "commodities continue to provide benefits to equity investors in terms of portfolio diversification". Gorton and Rouwenhorst (2006) and Erb and Campbell (2006) found that commodity prices show minor co-movements with equity markets and between each other prior 2001 (in some cases commodities and stock returns result negatively correlated). A second group led by the scientific work of Tang and Xiong (2010) and Silvennoinen and Thorp (2012), focuses on possible co-integration existing between the various markets and tries to verify the hypothesis of financialization of commodities before and after 2008.

Silvennoinen and Thorp (2012) introduce the DCC structure with an explicit treatment of expected stock volatility to analyze the integration between commodity and stock markets and the financialization of commodities. The authors conclude that there is an increasing correlation between the two markets in the period 1990–2009 (from 0 to 0.5) in US, EU and Japan.

The contribution of our paper with respect to this wide literature is twofold, first to address the issue of the most adequate correlation estimators among different markets, and second, to show that some known correlations between specific assets, as natural gas and crude oil, or different markets, as the stock and the commodity markets, no longer exist.

4.3 The Correlation Measure

Asset prices and asset returns show a time varying dynamics. In particular, most security prices result non stationary in mean and variance while returns result stationary only in mean. In presence of time varying volatility the measure of possible correlations between different securities represent a challenging issue. To measure the correlation existing between non stationary variables has been a common task among academics and practitioners. Time varying measures as HRC, ESC or the multivariate generalized autoregressive conditional heteroskedasticity

(MGARCH) method to generate the DCC estimator provide adequate measures in these cases. In most cases, the applications are made to the bivariate case. The Pearson product-moment correlation coefficient is the most used estimator to determine the strength and the direction between two random variables. A time varying unconditional correlation between two random variables r_1 and r_2 which have zero mean is defined by:

- $$\rho_{1,2;t} = \frac{E_{t-1}(r_{1,t}, r_{2,t})}{\sqrt{E_{t-1}r_{1,t}^2 E_{t-1}r_{2,t}^2}} \tag{4.1}$$

  where E_{t-1} is the expectation based on the observation available at time $t-1$. The conditional correlation is based on the information regarding previous periods.
- A dynamic correlation estimator is the HRC which applies (4.1) to a n-days rolling window. It gives equal weights to the n data of the interval and zero weights to previous observations:

$$\hat{\rho}_{1,2;t}^{HRC} = \frac{\Sigma_{s=t-n-1}^{t-1} r_{1,s}, r_{2,s}}{\sqrt{\Sigma_{s=t-n-1}^{t-1} r_{1,s}^2 \Sigma_{s=t-n-1}^{t-1} r_{2,s}^2}} \tag{4.2}$$

- The Exponential Smoother Correlation estimator, introduced by Risk Metrics to measure correlations between financial assets uses declining weights based on a parameter λ, which emphasizes current data but has no fixed termination point in the past where data becomes uninformative.

$$\tilde{\rho}_{1,2;t}^{ES} = \frac{\Sigma_{s=1}^{t-1} \lambda^{t-s-1} r_{1,s}, r_{2,s}}{\sqrt{(\Sigma_{s=1}^{t-1} \lambda^{t-s-1} r_{1,s}^2)(\Sigma_{s=1}^{t-1} \lambda^{t-s-1} r_{2,s}^2)}} \tag{4.3}$$

  The λ parameter allows to attribute differet weights to the various data. A crucial issue in using this estimator is the choice of the parameter λ. For λ very close to one this estimator is very close to the Pearson product-moment. Risk Metrics set $\lambda = 0.94$, values of $\lambda < 0.94$ provide values of this estimator less stable.
- The Dynamic Conditional Correlation is a natural extension of the GARCH models. The relationship between conditional correlations and conditional variance is obtained expressing the returns, $r_{i,t}$ as the conditional standard deviation times the standardized disturbance $\varepsilon_{i,t}$

$$h_{i,t} = E_{t-1} r_{1,t}^2 \qquad r_{i,t} = \varepsilon_{i,t}\sqrt{h_{i,t}} \tag{4.4}$$

The standardized disturbance $\varepsilon_{i,t}$ or the returns $r_{i,t}$ can be used in Eq. (4.2) to get:

$$\rho_{1,2;t}^{DCC} = \frac{\Sigma_{s=1}^{t-1} \varepsilon_{1,s}, \varepsilon_{2,s}}{\sqrt{\Sigma_{s=1}^{t-1} \varepsilon_{1,s}^2 \Sigma_{s=1}^{t-1} \varepsilon_{2,s}^2}} \tag{4.5}$$

and the correlation coefficient is given by:

$$\rho_{i,j,t} = \frac{q_{i,j,t}}{\sqrt{q_{i,t} q_{j,t}}} \tag{4.6}$$

Engle suggests that the $q_{i,j,t}$ process can be generated by a GARCH(1,1) model:

$$q_{i,j,t} = \bar{\rho}_{i,j} + \alpha(\varepsilon_{i,t-1}\varepsilon_{j,t-1} - \bar{\rho}_{i,j}) + \beta(q_{i,j,t-1} - \bar{\rho}_{i,j}) \tag{4.7}$$

Where: $\bar{\rho}_{i,j}$ is the unconditional correlation between $\varepsilon_{i,t}$ and $\varepsilon_{j,t}$.

4.3.1 Data Simulation

To compare the various estimators we generate a set of simulated data with a specific correlation structure. We choose four different correlation structures (functions) for the simulated returns.

The returns are generated using two Gaussian GARCH models, $h_{i,t}; i = 1, 2$, assuming (1) high or (2) low persistence:

$$h_{1,t} = 0.02 + 0.04 r_{1,t}^2 + 0.9 h_{1,t-1} \tag{4.8}$$

$$h_{2,t} = 0.4 + 0.4 r_{2,t}^2 + 0.4 h_{2,t-1} \tag{4.9}$$

$$\begin{Bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{Bmatrix} \sim N\left[0, \begin{pmatrix} 0.25 & \rho_t \\ \rho_t & 0.25 \end{pmatrix}\right]$$

$$r_{1,t} = \varepsilon_{1,t}\sqrt{h_{1,t}} \qquad r_{2,t} = \varepsilon_{2,t}\sqrt{h_{2,t}}$$

We select four functions to build different set of correlation structures:

1. Constant

$$\rho_t = 0.5 \tag{4.10}$$

2. Sine + Fast Sine

$$\rho_t = \begin{cases} \rho_t = 0.4 + 0.3\cos(\frac{2\pi t}{250}) & 0 < t < 700 \\ \rho_t = 0.4 + 0.3\cos(\frac{2\pi t}{50}) & 701 < t < 1300 \\ \rho_t = 0.4 + 0.3\cos(\frac{2\pi t}{250}) & 1301 < t < 2000 \end{cases} \tag{4.11}$$

3. Constant + Log.

$$\rho_t = \begin{cases} \rho_t = 0.1 & 0 < t < 500 \\ \rho_t = 25\frac{log(\frac{t}{2.5})}{log(t)} + 21.21 & 501 < t < 2000 \end{cases} \tag{4.12}$$

4. Two Step

$$\rho_t = \begin{cases} 0.3 & 0 < t < 700 \\ 0.8 & 701 < t < 1300 \\ 0.5 & 1301 < t < 2000 \end{cases} \tag{4.13}$$

- The first function shows a static correlation structure.
- The second function is built to take into account extreme events (crises)and allows a cyclical (seasonal) correlation where for $701 \leq t \leq 1300$ changes in the correlation structure are more frequent when an extreme event for $t > 1301$ occurs, correlations return to their previous structure.
- The third function considers a structural change in the correlation dynamics from poorly correlated to strongly correlated.
- The fourth correlation function accounts for strong structural changes. The occurrence of an extreme event causes an increase in correlation and a structural change.

We generate a 2×2000 matrix of observations for each function where the columns represent the time and the rows represent the returns. We measure correlation on this data using different estimators: The Pearson product-moment correlation coefficient, ρ; the HRC, $\hat{\rho}$, on different rolling windows: $50 - HRC$, $100 - HRC$, $250 - HRC$; the ESC, $\tilde{\rho}$[3]; and the DCC, ρ^{DCC}.

Four error metrics are used to assess the estimator's accuracy:

- *The Mean Square Error* (MSE) is the second moment of the error and incorporates the variance of the estimator and its bias. MSE enhances the large differences, between performed variables and desired variables:

$$MSE = \frac{\Sigma_{t=1}^{T}(\rho_t - \hat{\rho}_t)^2}{T} \tag{4.14}$$

[3]We use a λ coefficient of 0.94 which is calibrated from Risk Metrics.

Where ρ_t is the function described by (4.10), (4.11), (4.12), and (4.13), $\hat{\rho}_t$ is the correlation measured on the simulated data.

- The *Mean Absolute Error* (MAE) is the average of the absolute values of the differences between predicted variables and desired (true) variables:

$$MAE = \frac{\Sigma_{t=1}^{T}(|\rho_t - \hat{\rho}_t|}{T} \tag{4.15}$$

- The *Mean Absolute Percentage Error* (MAPE) is

$$MAPE = \frac{100}{T} \cdot \frac{\Sigma_{t=1}^{T}(\rho_t - \hat{\rho}_t)}{\Sigma_{t=1}^{T}|\rho_t|} \tag{4.16}$$

expresses the value of error relative to the actual value for observation t. The concept of MAPE is simple and convincing but it has limitations in practical application, like the possibility of zero value for actual observation results in division by zero. MAPE is zero in case of a perfect fit. Large errors can unfairly skew the overall error.

- The *Normalized Mean Square Error* (NMSE)

$$NMSE = \frac{\Sigma_{t=1}^{T}(\rho_t - \hat{\rho}_t)}{\Sigma_{t=1}^{T}(\rho_t - \bar{\rho}_t)} \tag{4.17}$$

generally shows the most striking differences among models. If a model has a very low NMSE, then it is well performing both in space and time. High NMSE values do not necessarily mean that a model is completely wrong. That case could be due to time and/or space shifting.

4.3.2 Comparing the Correlation Estimators

The four functions described in Eqs. (4.10), (4.11), (4.12), and (4.13) are reported in the plots of Fig. 4.1.

Each figure refers to the simulated series chosen and the various correlations estimators are compared for each series. The black line measure the effective correlation built for the specific simulated series. The various kind of correlation estimators adapt differently for each specific function, for instance in the case of the logarithmic function (Fig. 4.1c) the DCC seems to be the most accurate estimators, while the ESC the least accurate. To choose the most adequate estimator we compare different error metrics for each simulated series and report the values in Tables 4.1, 4.2, 4.3, and 4.4.

In Table 4.1 the results for the constant correlation structure are reported and the Pearson constant coefficient results the best estimator given the non time varying nature of the generated series.

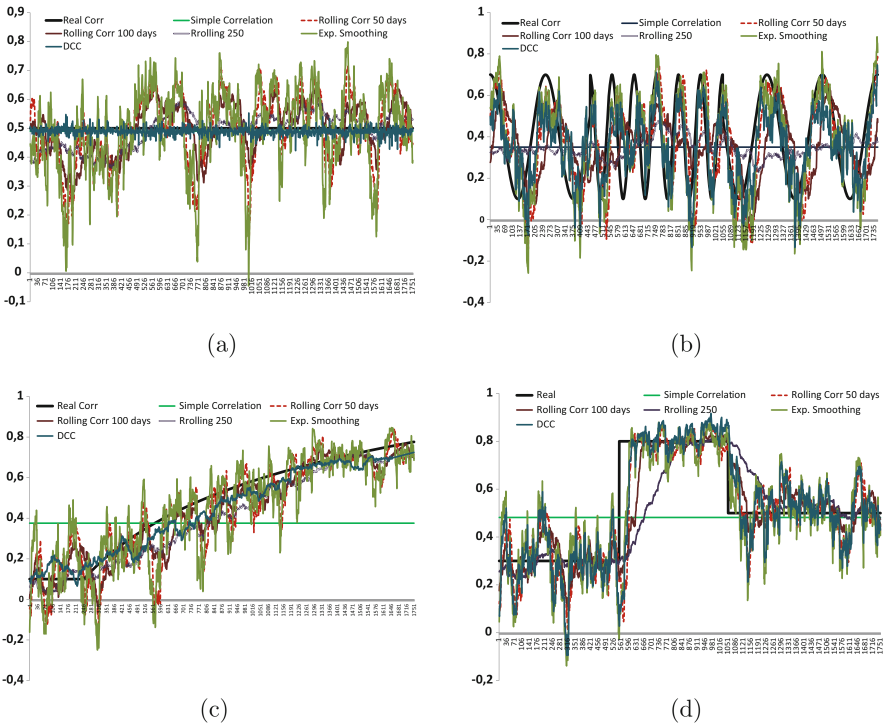


Fig. 4.1 Estimated correlation functions compared to the real correlation functions. (**a**) Constant function. (**b**) Sine +Fast Sine function. (**c**) Logarithmic function. (**d**) Two Step function

Table 4.1 Results for the constant correlation function

Error metric	ρ	HRC50	HRC100	HRC250	ESC	DCC
MSE	**0,01%**	1,42%	0,75%	0,36%	6,53%	0,02%
MAPE	**2,05%**	19,15%	14,74%	9,95%	41,94%	2,50%
MAE	**1,03%**	9,58%	7,37%	4,98%	20,97%	1,25%
NMSE	inf	inf	inf	inf	inf	inf

In Tables 4.2 and 4.3 the results for the monotonous simulated functions are reported. The DCC performs quite well in these cases and outperforms all the other estimators.

In Table 4.4 the results for the simulation with structural breaks are reported and in this case the $100 - HRC$ performs better than the other estimators. Given the various error metrics it is possible to rank the various estimators in Table 4.5. The simple Pearson product-moment results the best estimator only in the case of constant correlation between the series. In case of monotonic time varying series the DCC estimators results the most accurate while in the case of series showing structural breaks the HRC results the most accurate. This suggests the use of

Table 4.2 Results for the Sine + Fast Sine correlation function

Error metric	ρ	HRC50	HRC100	HRC250	ESC	DCC
MSE	4,559%	4,54%	5,72%	4,72%	3,17%	**2,49%**
MAPE	73,72%	69,25%	80,96%	72,64%	58,73%	**52,28%**
MAE	18,89%	17,71%	20,95%	18,99%	14,36%	**13,16%**
NMSE	1,031	1,026	1,295	1,067	0,718	**0,564**

Table 4.3 Results for the logarithmic correlation function

Error metric	ρ	HRC50	HRC100	HRC250	ESC	DCC
MSE	6,14%	1,33%	0,70%	0,79%	1,82%	**0,35%**
MAPE	76,06%	33,08%	24,13%	21,59%	38,66%	**16,02%**
MAE	22,20%	8,38%	6,59%	7,69%	9,83%	**5,01%**
NMSE	1,19	0,26	0,14	0,15	0,35	**0,07**

Table 4.4 Results for the Two Step correlation function

Error metric	ρ	HRC50	HRC100	HRC250	ESC	DCC
MSE	3,95%	1,71%	**1,34%**	2,06%	1,91%	1,79%
MAPE	31,88%	17,88%	**13,62%**	13,98%	21,34%	19,00%
MAE	15,54%	8,04%	**6,64%**	8,06%	9,24%	8,57%
NMSE	1,04	0,45	**0,36**	0,54	0,51	0,47

Table 4.5 Ranking the estimators

Rank	Constant	Sine + Fast sine	Logarithmic	Two step
1	Simple correlation	DCC	DCC	HRC 100
2	DCC	Exp.smoother	Exp. smoother	HRC 50
3	HRC 250	HRC 50	HRC 100	HRC 250
4	HRC 100	HRC 250	HRC 250	DCC
5	HRC 50	HRC 100	HRC 50	Exp. smoother
6	Exp. Sm.	Simple corr.	Simple corr.	Simple corr.

different estimators in presence of different structures of the data. The accuracy of the HRC estimators depends on the size of the rolling windows. If there exist a very dynamic structure with frequent changes, it is convenient to chose a small rolling window, for more persistent structures a large window should be chosen instead. In principle the size of the window should be set as the result of an optimization problem aimed to find the size n which minimizes the volatility of the HRC coefficient. This problem is not addressed here and it is discussed in D'Ecclesia and Jotanovic (2015).

The Exponential Smoother Correlation is a particular case of the DCC, when the parameter $\alpha + \beta = \lambda$. It can capture the dynamics when frequent and strong changes occur but it is very unstable when a persistent correlation structure exist. DCC get unstable for strong structural changes, it is stable for persistent correlation structures.

4.4 Measuring Correlation

We study the changing relationship existing between commoditiies and financial indices using the HRC and the DCC estimators. The following set of data[4] are used:

1. Energy
   (a) WTI Future Nymex (CL1) in USD/bbl
   (b) Brent ICE (CO1) in USD/bbl.
   (c) US Gas Henry Hub (NG) in USD/MMbtu
   (d) UK NBP Gas (FN1) in GBP/therm
2. Indexes
   (a) DJ UBS – Commodity index TR – Return ind. (DJ-UBS).
   (b) MSCI AC WORLD USD(MSACWI) – Price index.
   (c) Morningstar Global Government Bond Index (MSBI).

The data used in the research are daily data from March 2000 to December 2014.[5] HRC or DCC methods are used to measure the corelations existing between the various sectors or assets.

4.4.1 Stationarity of the Series

We first test each price series for stationarity, the Augmented Dickey Fuller test shows that all the price series have a unit root and therefore are $I(1)$ variables. The ADF test rejects the null Hypothesis H_0 of unit roots for the log-returns which all result $I(0)$. Engle's test is used to verify the presence of heteroskedasticity, the test allows to reject the null Hypothesis in favour of the presence of heteroskedasticity. We do not report the results of the test here, they are available upon request.

4.4.2 Structural Breaks

We further investigate the occurrence of structural breaks in the price series with the aim to analyse the correlation dynamics. The dynamics of the MS Gov Bond Index (MSBI), the Stock Index (MSACWI) and the Commodity Index (DJ-UBS Index) are reported in Fig. 4.2.

We use the BAI and Perron (1998) test for multiple structural changes (or breaks). The test identifies four structural breaks for almost all the variables. In Table 4.6 the results of the Bai Perron test for structural breaks are reported for each series.

---

[4]The database is obtained from Bloomberg and Data Streaming data platforms.

[5]Except for Morningstar GG Index, which was constructed in 2001, and for energy commodities data that are available from March 2000 to January 2014.

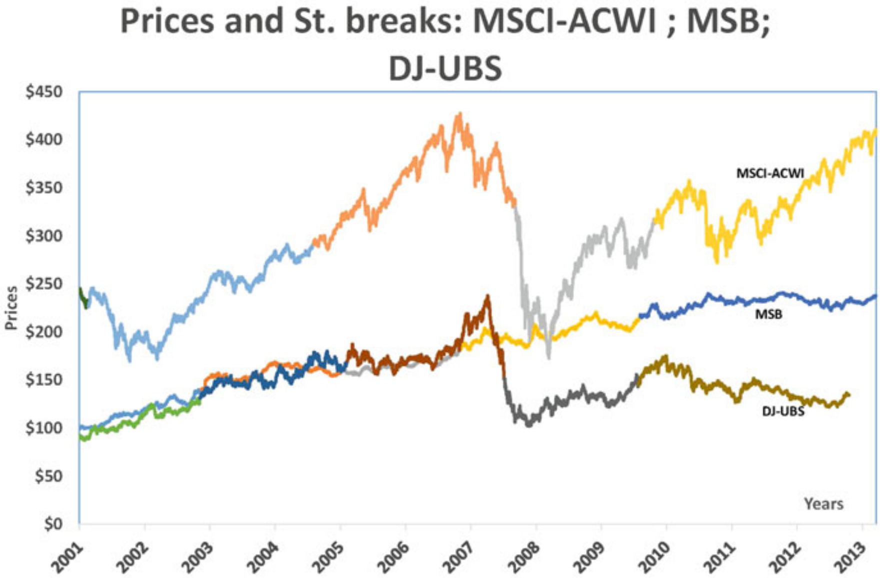


Fig. 4.2 Commodity, stock and bond indices (20/3/2000–21/1/2014)

Table 4.6 Structural breaks for the market indices and the energy commodities

Variables	**nr**	Date	Date	Date	Date
MSCI-AWCI	4	18/02/2002	01/08/2005	02/09/2008	19/10/2010
MS BI	4	13/11/2003	04/01/2006	02/11/2007	02/08/2010
UBS J COMM	4	01/12/2003	11/04/2006	06/10/2008	17/11/2010
WTI	4	29/01/2003	20/06/2005	30/08/2007	01/12/2010
Brent	4	17/11/2003	03/01/2006	14/10/2008	21/12/2010
HH	4	22/10/2002	20/10/2004	20/01/2009	21/04/2011
NBP	4	16/07/2003	16/09/2005	17/10/2008	26/11/2010

The MSBI shows a constant uptrend for the entire period while MSCI ACWI is more volatile and shows the presence of structural breaks and a strong downtrend during the crisis periods. For this reason the log returns correlation between the two indexes is cyclical. The structural breaks for MSCI-ACWI Stock index are identified on January 2002, June 2005, September 2008 (when the crisis period starts) and December 2010 right before the downgrading of the US credit rate. In the case of the DJ-UBS Commodity index the first structural break occurs in 2003. The other are found in 2006, 2008 and in 2010.

The energy commodity prices have the same number of structural breaks and they are identified for crude oil in 2002, 2008, 2010 and 2012. For the natural Gas contract the most important structural break occurred in 2005–2006. The second in 2010, the others in 2003 and 2008–2009. A more accurate interpretation of the various breaks see Dec Jotanovic (2015).

4.4.3 Correlations Between Commodities and Financial Markets

We use the DCC approach to estimate the correlation between the commodity Index and the stock index during the period 2000–2014. In Table 4.7 the main statistics (μ,σ, max, min) for the estimated conditional correlations are reported.

The correlation between the stock (MSCI-AW) and the commodity indices (DJUBS) is important to investigate. The two indices show opposite dynamics over the period 2000–2003 then they seem to have a common behavior until 2006 to diverge again until 2008, when the big crash of 2008 occurs and generates a definitively different pattern over the last 6 years. A changing pattern can be detected in the correlation dynamics: (i) a correlation coefficient swinging between −0.3 to +0.3 in the period 2000–2003, (ii) a prevailing positive correlation for the following 5 years until the explosion of the GFC in 2008; (iii) a drop in correlation to −0.1 in september 2008 to increase after 2 months to a positive +0.4 and to stay above 0.4 until 2011. After that a new pattern can be identified.

Correlation between the MSCI-AW and the MSGBI fluctuates in the range −0.4, +0.4 with period of negative correlations (2000–2003), (2006–2008) alternating to period of positive correlations. The period of negative correlations is much shorter compared to those of positive correlations, even if it fluctuates between 0 and +0.4. The dynamnics of the two indices show how in periods of economic recession the dynamics of bond and stock prices behave differently: stock prices decline, interest rates are lowered causing increases in bond prices.

Correlation between MSGBI and DJ UBS fluctuates between 0 and 0.3 showing a low positive correlation over the entire period. The volatility of the correlation results higher before the 2008 GFC compared to the post crisis period. It is interesting to see that Commodity markets and Government Bond markets seem to be poorly correlated and with a stable performance.

The analysis of the correlations dynamics could provide additional insights for the understanding of each market behavior so we test for structural breaks in the correlation series. Using the Bai-Perron test on the series of the correlation coefficients we are able to find 4 structural breaks for each estimated correlation, Figs. 4.3, 4.4, and 4.5.

The first structural change is identified at the end of 2003 which is the period when most of scholars identify with the beginning of the financialization process for the commodity markets. During the crisis the correlations between stock and

Table 4.7 HRC and DCC descriptive statistics

Assets	μ	σ	Min	Max
DCC(DJUBS,MSCI-AW)	0.40	0.234	−0.38	0.62
DCC(DJUBS, MSGBI)	0.15	0.10	−0.02	0.32
DCC(MSCI-AW,MSGBI)	0.12	0.40	−0.48	0.41
HRC(WTI,HH)	0.10	0.01	−0.01	0.01
HRC(Brent,NBP)	0.02	0.10	−0.01	0.05

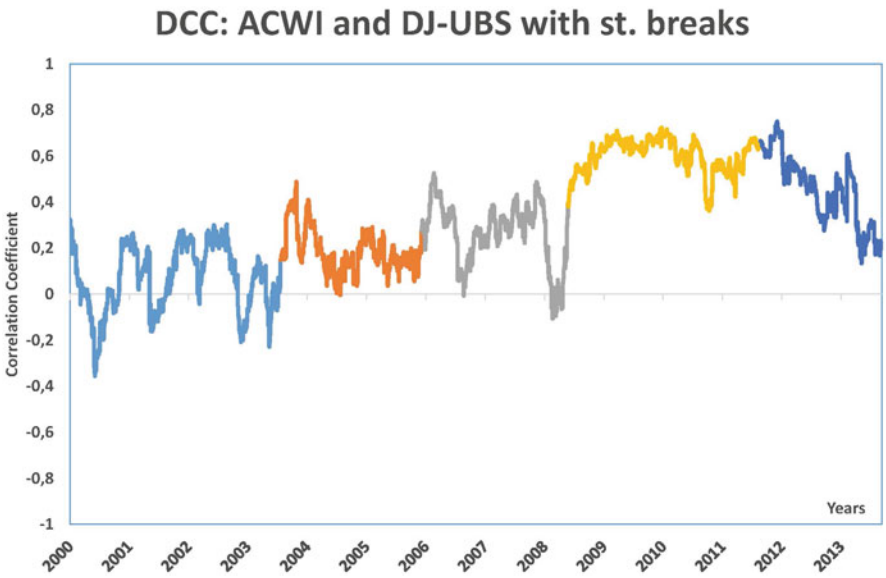


Fig. 4.3 Structural breaks on correlation function between ACWI and DJ-UBS (20/3/2000–21/12/2014)

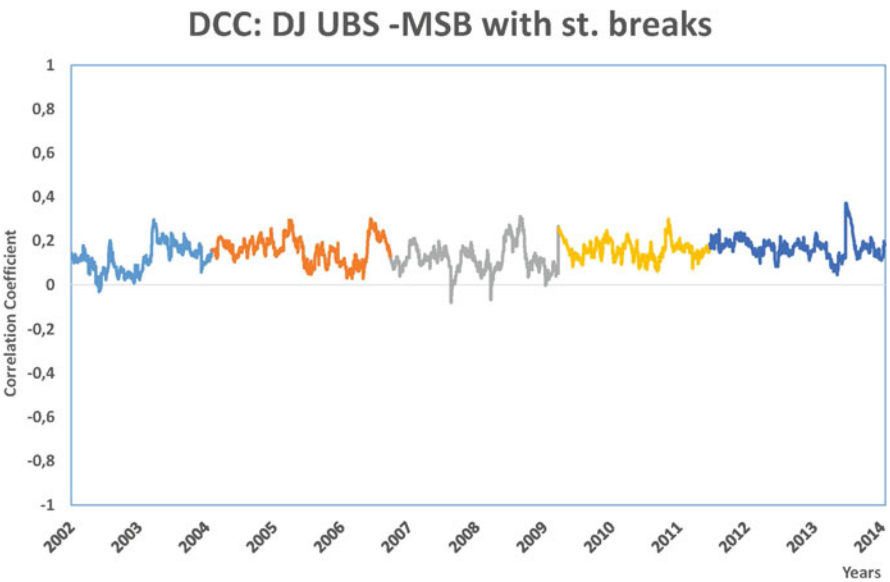


Fig. 4.4 DCC correlations between MSGBI and DJUBS (20/3/2000–21/1/2014)

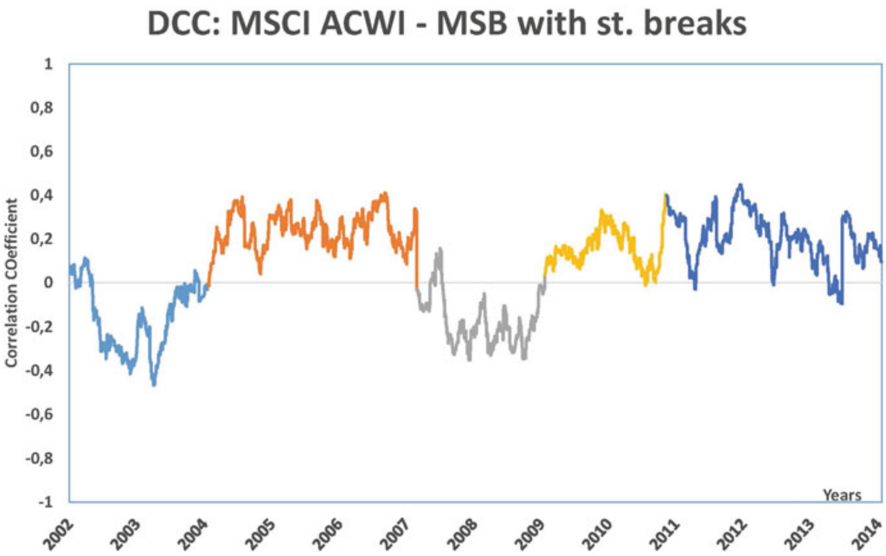


Fig. 4.5 DCC correlations between MSGBI and MSCI-AW (20/3/2000–21/1/2014)

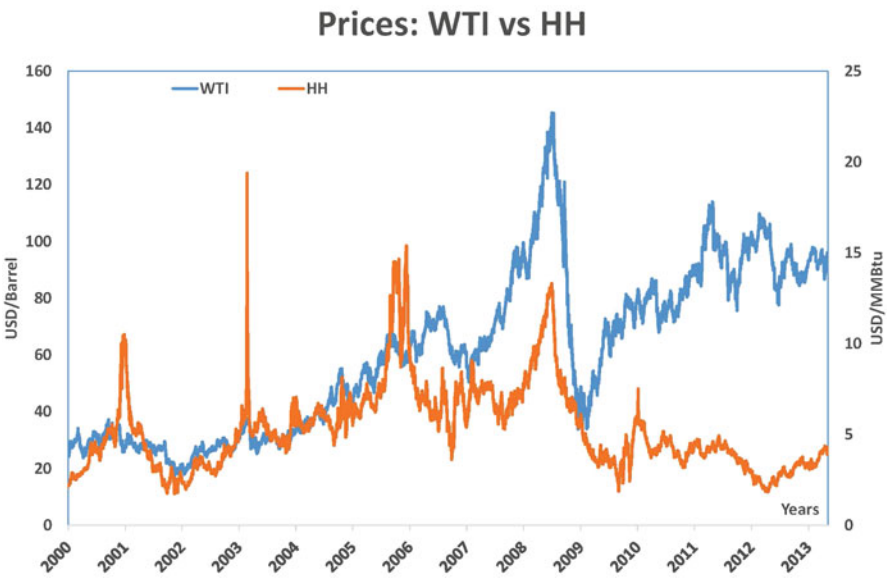


Fig. 4.6 WTI and HH daily prices (20/3/2000–21/1/2014)

commodities markets increased (more than doubled) reaching the highest values, almost 0.7. Before the crisis, there is a return to zero of the correlation structure. This is due to the fact that stock prices started to drop 6 months earlier that the commodity index (Fig. 4.2).

4.4.4 Correlations Between Energy Commodities

Crude oil and natural gas prices have experienced very volatile dynamics due to changing features of their fundamentals (supply and demand) over the last decade. Figures 4.6 and 4.7 illustrate the dynamics of crude oil and natural gas markets over the period 2000–2014 in the US and European markets. They show how the two markets have experienced large periods of fluctuations due to structural changes in the supply and demand. The crude oil markets in the past decade have been affected by various events.

- a renewed solidarity within OPEC
- OPEC new relationship with Russia
- China accounting for 38% of Global GDP growth
- India becoming a major importer, refiner and re-exporters of products
- the impending Chinese crude oi futures contract portend major new structural changes

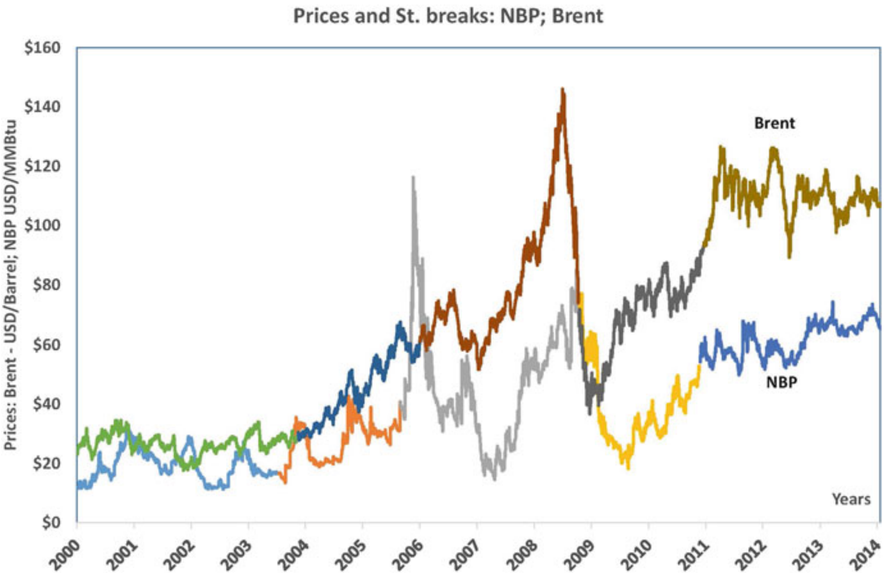


Fig. 4.7 Brent and NBP daily prices (20/3/2000–21/1/2014)

The natural gas market, as well has witnessed major changes:

- the increasing production of unconventional gas (shale Gas) in the US started in 2009 as the EU's domestic production peaked
- the shift toward US and Australian gas exports from Qatari dominance in Liquefied NG
- the de-coupling of NG contracts from oil based formulas
- the development of new efficient combined cycle gas turbines with low level of pollution emissions
- the European Community third Gas Package fostering the creation of a single market for NG

this is only to cite some of the most important events which have affected the Oil and Natural Gas market.

For a more accurate understanding of the recent dynamics of crude oil and natural gas prices see Kilian (2009), Kilian and Murphy (2014), Pindyck (2004) and Rogers (2010) only to cite some of the most relevant contributions on the topic.

Figure 4.8 shows the correlation dynamics between the Brent-crude oil log returns and the NBP-Natural gas log returns. The $HRC - 100$ is estimated and two commodities result almost no correlated, with a correlation coefficient which fluctutaes between 0.05 and 0.10. The US natural gas market, HH, shows a statistically significant, positive, almost constant (in average 0.13) correlation with the crude oil market (WTI) (see Fig. 4.9).

Some structural changes are also found in the correlation pattern, showing a stable period in the first 5 year (2000–2005), a changing pattern with a slightly

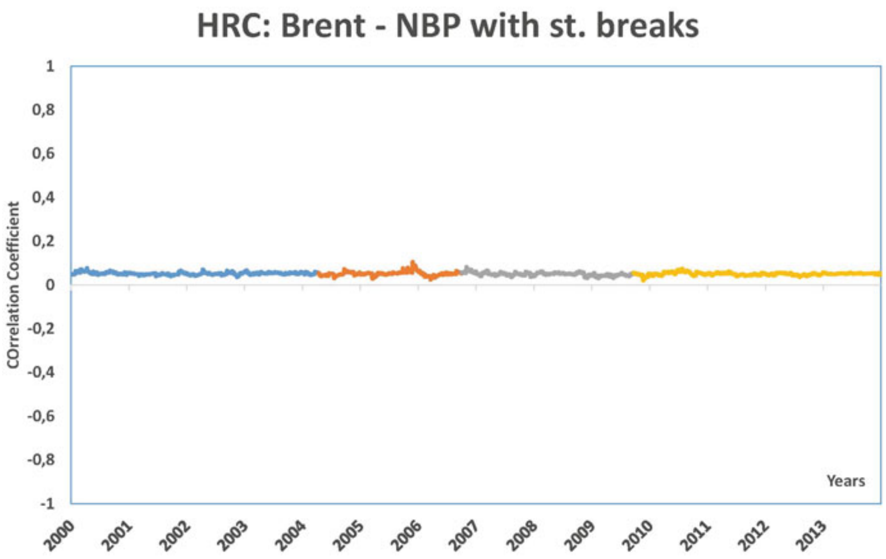


Fig. 4.8 HRC between Brent and NBP (20/3/2000–21/1/2014)

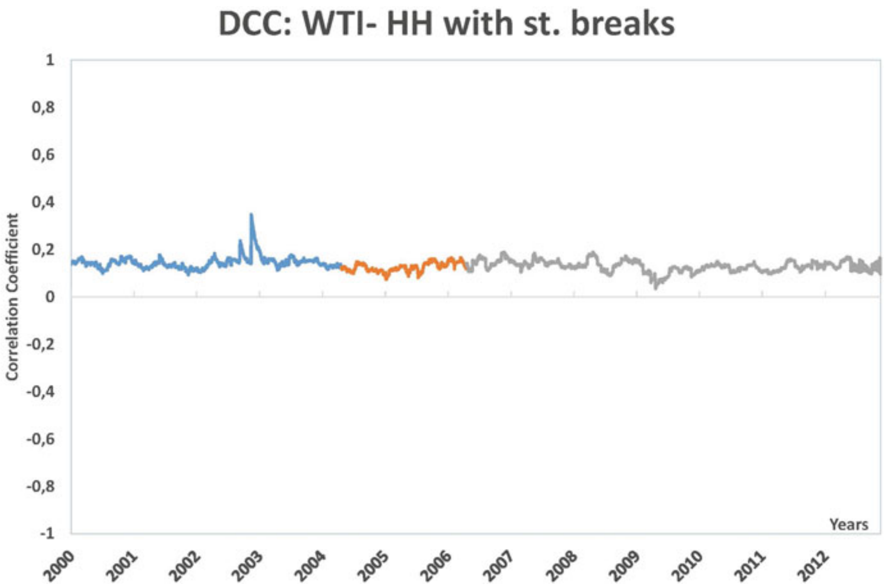


Fig. 4.9 100-HRC between WTI and HH log returns (20/3/2000–21/1/2014)

increasing coefficient in the period 2005–2007 which is the period of large increase of the supply of natural gas in Europe and an increasing volatility of oil prices, see Fig. 4.9.

4.5 Concluding Remarks

Measuring correlations represents a key task for risk management and derivative pricing in financial markets. We have shown how to identify the most adequate correlation estimator in presence of time varying and heteroschedastic returns of commodities and financial assets. HRC with different size rolling windows and the DCC estimators provide accurate measures in presence, respectively, of structural breaks or heteroskedasticity in the log returns. The DCC correctly captures the correlations existing between the stock market, the commodity markets and the bond market; while the HRC-100 results adequate to measure the correlation between two energy commodities as the crude oil and the natural gas. The commodity market has modified its role respect to the stock market becoming more integrated with it after the Global Financial Crisis, while no changes can be identified respect to the bond market. This allows fund managers and investors to update their investment strategies or portfolios selections.

The relationship existing between two major energy commodities: crude oil and natural gas in Europe and in the US has been changing dramatically. There is a common belief that these energy sources are correlated with each other. We found very low or no correlation at all between the two fossil fuels in both the European and the US market showing how recent changes in the supply and demand of these markets have modified a well known positive relationship existing between the two commodities.

References

J. Bai, P. Perron, Estimating and testing linear models with multiple structural changes. Econometrica **66**, 47–78 (1998)

T. Bollerslev, Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. Rev. Econ. Stat. **72**, 498–505 (1990)

T. Bollerslev, R. Engel, J. Woldridge, A capital asset pricing model with time varying covariances. J. Polit. Econ. **96**, 116–31 (1988)

T. Bollerslev, R. Engle, D. Nelson, Arch models. Handb. Econ. (1994)

B. Büyükşahin, M. Haigh, M. Robe, *Commodities and Equities: A "market of one"?* **4**, 2959–3038 (Nova Science Publication, New York, 2009)

R.L. D'Ecclesia, V. Jotanovic, Correlations in the natural gas market. PhD School of Economics Working Papers (2015)

Z. Ding, R. Engle, Large scale conditional matrix modeling, estimation and testing'. Academia Economic Papers (2001)

R. Engle, Dynamic conditional correlation – a simple class of multivariate garch models. J. Bus. Econ. Stat. **20**(3), 339–350 (2002)

R. Engle, J. Mezrich, Garch for groups. Risk Aug. **9**, 36–40 (1996)

R. Engle, N.G. Victor, M. Rothschild, Asset pricing with an arch covariance structure empirical estimates for treasury bills. J. Econ. **45**, 213–237 (1990)

C. Erb, H. Campbell, The strategic and tactical value of commodity futures. Financ. Anal. J. **62**, 69–97 (2006)

G. Gorton, G. Rouwenhorst, Facts and fantasies about commodity futures. Financ. Anal. J. **62**, 47–68 (2006)

G. Gorton, F. Hayashi, G. Rouwenhorst, The fundamentals of commodity futures returns. Working paper, Yale University (2007)

L. Kilian, Not all oil price shocks are alike: disentangling demand and supply shocks in the crude oil market. Am. Econ. Rev. **99**(3), 1053–1069 (2009)

L. Kilian, D. Murphy, The role of inventories and speculative trading in the global market for crude oil. J. Appl. Econ. **29**, 454–478 (2014)

R.S. Pindyck, Volatility in natural gas and oil markets. J. Energy Dev. **30**, 1–19 (2004)

H. Rogers, Lng trade-flows in the Atlantic basin: trends and discontinuities (The Oxford Institute for Energy Studies, Oxford, 2010)

A. Silvennoinen, S. Thorp, Financialization, crisis and commodity correlation dynamics. J. Int. Financ. Mark. Inst. Money **24**, 42–65 (2012)

K. Tang, W. Xiong, Index investment and financialization of the commodities. NBER working paper series (2010)

Chapter 5
Measuring *Model Risk* in the *European Energy Exchange*

Angelica Gianfreda and Giacomo Scandolo

Abstract It has been shown that model risk has an important effect on any risk measurement procedures, hence its proper quantification is becoming crucial especially in energy markets, where market participants face several kinds of risks (such as volumetric, liquidity, and operational risk). Therefore, relaxing the assumption of normality and using a wide range of alternative distributions, we quantify the *model risk* in the German wholesale electricity market (the *European Energy Exchange*, EEX) by studying day–ahead electricity prices from 2001 to 2013 using the well-established setting of GARCH–type models. Taking advantage of this long price history, we investigate the "time evolution" of the measured model risk across years by adopting a rolling window procedure. Our results confirm that the increasing complexity of energy markets has affected the stochastic nature of electricity prices which have become progressively less normal through years, hence resulting in an increased *model risk*.

Keywords VaR • Risk measures • Electricity market • Spot and day–ahead prices • Germany • RES

5.1 Introduction and Background

It has been shown that model risk has an important effect on any risk measurement procedures, hence its proper quantification is becoming crucial especially in energy markets, where traders and market participants face several kinds of risks such as volumetric, liquidity, and, more importantly, operational risk. Therefore, we propose for the first time the assessment of model risk in the German wholesale electricity market, looking at daily spot prices and comparing several models presented in the

A. Gianfreda (✉)
DEMS Department, University of Milano-Bicocca, Milan, Italy
e-mail: angelica.gianfreda@unimib.it

G. Scandolo
Department of Economics and Management, University of Firenze, Florence, Italy
e-mail: giacomo.scandolo@unifi.it

G. Consigli et al. (eds.), *Handbook of Recent Advances in Commodity and Financial Modeling*, International Series in Operations Research & Management Science 257,
DOI 10.1007/978-3-319-61320-8_5

literature with their possible variations. We adopt a quantitative measure of model risk when choosing a particular reference model within a given class: namely, the *relative measure of model risk*, as proposed by Barrieu and Scandolo (2015).

The specification of a model is a crucial step mainly based on the assumption on the reference empirical distribution. And, it has been observed that the final risk figure is sensitive to the choice of the model. Hence, working with a potentially not well–suited model is referred to as *model risk*. The study of the impact of model risk and its quantification is an important step in the whole risk measurement procedure, but we are not going into the issue of *model uncertainty*.

We quantify the *model risk* by studying day–ahead electricity prices in the *European Energy Exchange*, EEX. Germany, indeed, decided to exit from nuclear power by 2020 focussing on renewable energy sources and energy efficiency. This market is characterized by an high wind penetration which has increased the complexity of the electricity price dynamics given that wind energy (and solar) is highly variable and partially predictable. In this context, renewables accounted for more than 71% of total electricity capacity addictions in EU in 2011, bringing the share of renewable energy over total electric capacity to 31.1%; with Germany being the leading market (see REN21 2014). Given that the marginal cost of wind production is near zero, negative pricing has been introduced in some countries like Germany, which today is the market with the highest hourly frequency of observed negative prices. "Negative prices" are power wholesale market signals occurring when high inflexible power generation meets low demand. Power sources are considered inflexible when they cannot be shut down and restarted in a cost–efficient way. This has contributed to the profound structural changes that occurred in Germany over the last two decades; as for the introduction of emission trading, the nuclear power phase–out, and the growing electricity generation from renewable energy sources (RES), also promoted by the renewable energy sources act (*Erneuerbare-Energien-Gesetz*, EEG).

Model risk issues for energy markets have been recently presented by Bannör et al. (2016) who propose the quantification of parameter uncertainty in complex stochastic models, considering the sensitivities of a derivative value corresponding to a specific pricing model. This setting is used to quantify model risk in the economic evaluation of power plants. Therefore, they do not perform model analysis and comparisons across estimation methods, empirical distributions and consequent approximations of quantities; and, we aim at filling this gap assessing the model risk by computing a simple measure based on Value-at-Risk (VaR). VaR is commonly applied since it measures market risk by means of the probability distribution of a random variable and evaluates the risk with a single real number. As a result, VaR has become an easy, immediate and essential tool within financial markets. More specifically, VaR measures the worst expected loss under usual market conditions over a specific time interval at a given confidence level. As J.P. Morgan states *VaR answers the question: "how much can I lose with x% probability over a pre–set horizon?"*. In other words, VaR corresponds to a specific quantile (of order α) of the distribution of the portfolio losses over the horizon T. Therefore, T and α are the two main parameters that should be chosen in a way appropriate for the purpose of risk measurement: the time horizon T can differ from few hours (for an active trading desk) to a year (as for instance for a pension fund); whereas the quantile order α is typically very small (1% for regulatory requirements or 5% for company internal risk management). We look at the VaR measure by combining two well–known and deeply studied facts of dynamic or time–varying volatility and non–normality of price "changes" distributions. We are forced to consider price changes instead of logarithmic price returns since electricity prices can vanish or even become negative.

A widespread methodology for VaR assessment is the *variance–covariance* approach, according to which the portfolio VaR is computed under the assumption that asset returns are jointly normally distributed. The computation then relies on the historical estimation of the covariance matrix. An alternative methodology is the *historical simulation*, which does not make any distributional assumptions, and the quantiles are directly retrieved from the empirical distribution of past observed returns. Both methodologies prove flawed: the variance-covariance approach is based on the normality assumption, which is often a poor description of return distributions; and, the historical simulation, though dispensing with the normality hypothesis, is generally viewed as a too backward-looking measure, failing to capture conditional distributions of returns.

In order to overcome these problems, the above approaches have been modified and improved in various ways. For instance, weighted historical simulation allows to give more weight to the most recent observations. And, within the variance-covariance approach, different volatility dynamics specifications and/or alternative hypotheses on the conditional distributions of returns have been proposed, as for instance the multivariate student. These allow to better capture well-documented empirical facts such as dynamic volatilities[1] and fat tails. A non-exhaustive list in this vein includes: exponentially weighted moving average estimates of covariances (RiskMetrics approach), normal and student GARCH models, skewed student APARCH and skewed student ARCH models, historical simulation with autoregressive moving average forecasts (filtered historical simulation), extreme value theory estimates of quantiles (see Bhattacharyya and Ritolia, 2008, among others).

Having observed market structural changes occurred in Germany together with frequent negative values affecting the stochastic nature and properties of electricity prices, (Gianfreda and Bunn, 2015) provided evidence on empirical shapes different from the normal one, hence suggesting the use of different distributions to compute VaR measures more suitable for electricity markets. Given that the common assumption of standard normal distribution is still considered even if it represents an important shortcoming in existing VaR methods, we intend to quantify model risk and provide convincing empirical evidence when this normality assumption is relaxed.

[1]The volatility dynamics within energy markets has been extensively studied and modelled by using GARCH–type models, see for instance Fan et al. (2008) who measured risk for both the WTI and Brent crude oil spot markets using the *generalized error distribution* (GED) to estimate the extreme downside and upside VaR of oil price returns.

Our aim is to quantify the model risk in energy markets by using a measure proposed by Barrieu and Scandolo (2015). This type of figures can be useful to traders, when confronted with the choice of the model to be employed. The measure we use produces a pure number, independent from the reference currency, hence allowing for immediate comparisons across different countries. Moreover, it takes values in the range $[0, 1]$ and vanishes precisely when there is no model risk. It crucially depends on the choice of a *reference* distribution and a set of *alternative* distributions. Below, we will comment on the rationale behind our choices.

Our approach for assessing model risk is very general. We consider the well-known GARCH methodology and we compute the measure of model risk, based on both the worst– and best–case VaR under a wide range of alternative distributions for the innovations.

The paper is structured as follows: in Sect. 5.2 we recall the notation and the definition of the measure of model risk that we implement; Sect. 5.3 briefly describes the data used, and Sect. 5.4 presents the framework of the GARCH models for the empirical analysis, and the procedure to forecast quantiles. Numerical results are provided in Sect. 5.5, whereas Sect. 5.6 concludes discussing another methodology with great potential applications.

5.2 The Relative Measure of Model Risk

Let us recall now the definition of the *relative measure of model risk*, recently proposed by Barrieu and Scandolo (2015). This measure quantifies the model risk we are exposed to when estimating Value–at–Risk (VaR) using different alternative models. We will use a slightly different route to introduce this measure, perfectly equivalent to the original one, but better suited to our framework.

If X is a random variable defined on some probability space $(\Omega, \mathcal{F}, P)$, its cumulative distribution function (CDF) is $F_X(t) = P(X \leqslant t)$ and the (upper) quantile of order $\alpha \in (0, 1)$ is $q_\alpha(X) = \inf\{t : F_X(t) \geqslant \alpha\}$. When X is interpreted as a portfolio return or Profit-and-Loss variable, then the Value–at–Risk at level $\alpha \in (0, 1)$ is

$$\mathrm{VaR}_\alpha(X) = -q_\alpha(X)$$

in such a way that VaR_α is the maximum loss that can be suffered, once we exclude the α fraction of worst cases. Notice that $\mathrm{VaR}_\alpha(aX + b) = a\mathrm{VaR}_\alpha(X) - b$ always holds, with $a > 0$. Typical values for α are 5%, 1%, or even 0.1%.

If we consider another probability on $(\Omega, \mathcal{F})$, say Q, then the CDF of X will change to $F_{X,Q}(t) = Q(X \leqslant t)$. Accordingly, the quantile $q_{\alpha,Q}$ and the Value–at–Risk $\mathrm{VaR}_{\alpha,Q}$ will vary, depending on the choice of Q. Let $\mathcal{Q}$ be a set of probabilities on $(\Omega, \mathcal{F})$ and fix $Q_0 \in \mathcal{Q}$. The distribution of X under Q_0 acts as the *reference model*, while the distributions under the other $Q \in \mathcal{Q}$ are the *alternative models*. For a simple example, one could select X, $\mathcal{Q}$ and Q_0 is such a way that $X \sim N(0, 1)$ under Q_0 and X has alternative (standard) distributions under other probabilities $Q \in \mathcal{Q}$.

Once X, Q_0 and $\mathcal{Q}$ have been fixed, and for a given $\alpha \in (0, 1)$, we define

$$\underline{\mathrm{VaR}}_\alpha = \inf_{Q \in \mathcal{Q}} \mathrm{VaR}_{\alpha,Q}(X) \qquad \overline{\mathrm{VaR}}_\alpha = \sup_{Q \in \mathcal{Q}} \mathrm{VaR}_{\alpha,Q}(X) \tag{5.1}$$

assuming they are both finite, with $\underline{\mathrm{VaR}}_\alpha \neq \overline{\mathrm{VaR}}_\alpha$. Observe that the interval

$$\left[\underline{\mathrm{VaR}}_\alpha, \overline{\mathrm{VaR}}_\alpha\right]$$

can be interpreted as the range of possible values for VaR, under all the alternative models in $\mathcal{Q}$ that we are considering; although not all internal values of the range necessarily correspond to a given model (for instance, this occurs when $\mathcal{Q}$ is a *finite* set).

The *Relative Measure of Model Risk* (RMMR) for VaR_α associated to Q_0 and $\mathcal{Q}$ is then defined as

$$\mathrm{RMMR}(Q_0, \mathcal{Q}) = \frac{\overline{\mathrm{VaR}}_\alpha - \mathrm{VaR}_{\alpha,0}}{\overline{\mathrm{VaR}}_\alpha - \underline{\mathrm{VaR}}_\alpha} \tag{5.2}$$

where $\mathrm{VaR}_{\alpha,0} = \mathrm{VaR}_{\alpha,Q_0}(X)$ is the risk figure under the reference model. In what follows, we will drop the two arguments Q_0 and $\mathcal{Q}$ as they will be clear from the context. Notice that, since $\underline{\mathrm{VaR}}_\alpha \leqslant \mathrm{VaR}_{\alpha,0} \leqslant \overline{\mathrm{VaR}}_\alpha$, this measure takes values in $[0, 1]$ and this fact allows its use in various situations and across different markets. Also, we note that enlarging the set $\mathcal{Q}$ does not necessarily yield to a higher RMMR figure, as both the numerator and the denominator in (5.2) increase. In other words, considering additional alternative models need not increase model risk. For further properties and discussion about RMMR we refer to Barrieu and Scandolo (2015).

If we interpret the *reference distribution* as the model actually used in risk forecasting, the two extreme values for RMMR can be interpreted as follows:

- $\mathrm{RMMR} = 0$ whenever $\overline{\mathrm{VaR}}_\alpha = \mathrm{VaR}_{\alpha,0}$, i.e. the maximum VaR is exactly the one computed under the reference distribution; in this case there is ***no model risk***, since using any alternative model provides a lower risk figure;
- $\mathrm{RMMR} = 1$ whenever $\underline{\mathrm{VaR}}_\alpha = \mathrm{VaR}_{\alpha,0}$, i.e. any other assumptions apart the reference one produce higher VaR figures; in this case there is ***full model risk***.

In our empirical study, we consider RMMR with respect to three different *reference models*, by specifying three different (standard) distributions for the innovations of a suitable GARCH process. One of these reference models will be the standard normal, but then we will consider two other models allowing for asymmetry and thick tails. For all three choices of Q_0, we consider the same set $\mathcal{Q}$ of alternative models, including the three possible reference models and 7 additional models.

We stress again that the measure of model risk RMMR depends on Q_0 and $\mathcal{Q}$. Depending on which probability models are entered in $\mathcal{Q}$, the relative position of the reference model can in principle swing from being very conservative to being very optimistic.

Below, we will explain the rationale behind our choices for Q_0 and $\mathcal{Q}$. We will also provide evidence that, in a sense, the qualitative conclusions we obtain do not depend too much on the composition of the set $\mathcal{Q}$.

5.3 Data and Preliminary Analysis

Our data-set consists of 24 hourly price series determined on daily basis, from Mondays to Fridays and adjusted for holidays, as collected from Datastream from 01/01/2001 to 31/12/2013 (hence consisting of 3392 prices), denoted S_t^h, $h = 1, \ldots, 24$. Daily averages have been computed and they represent the daily 'spot' or 'day-ahead' electricity prices, $S_t = \frac{1}{24} \sum_h S_t^h$. The time series of price levels is depicted in Fig. 5.1, and it clearly shows that null or even negative electricity prices can be determined.[2] Among *stylized facts* for financial asset returns, we then confirm that German electricity price differences show little autocorrelation,

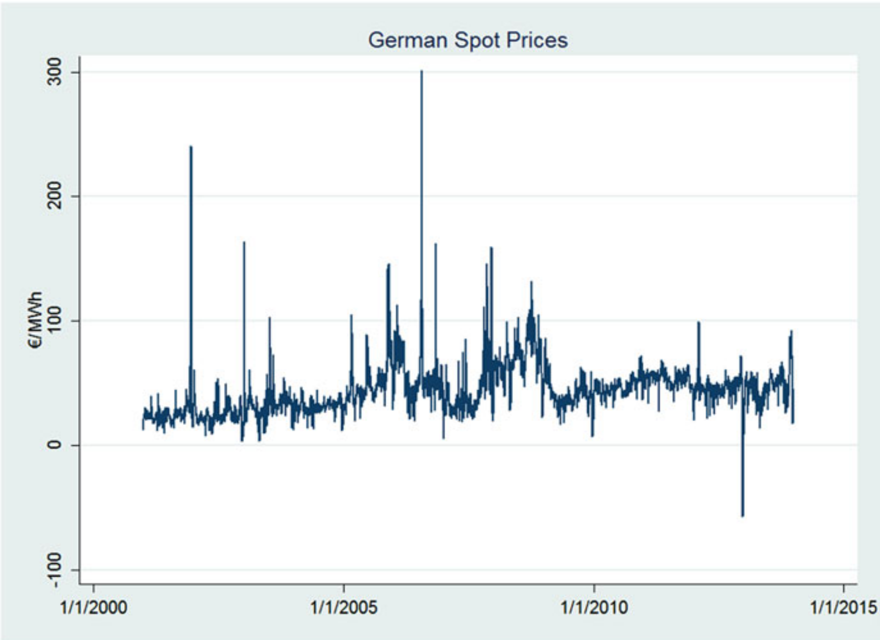


Fig. 5.1 Dynamics of "day–ahead" or "spot" daily electricity prices in Germany from 2001 to 2013

[2]This new empirical fact is even more evident when hourly prices are considered; and less so, when day-ahead prices are computed as arithmetic mean of 24 hourly prices.

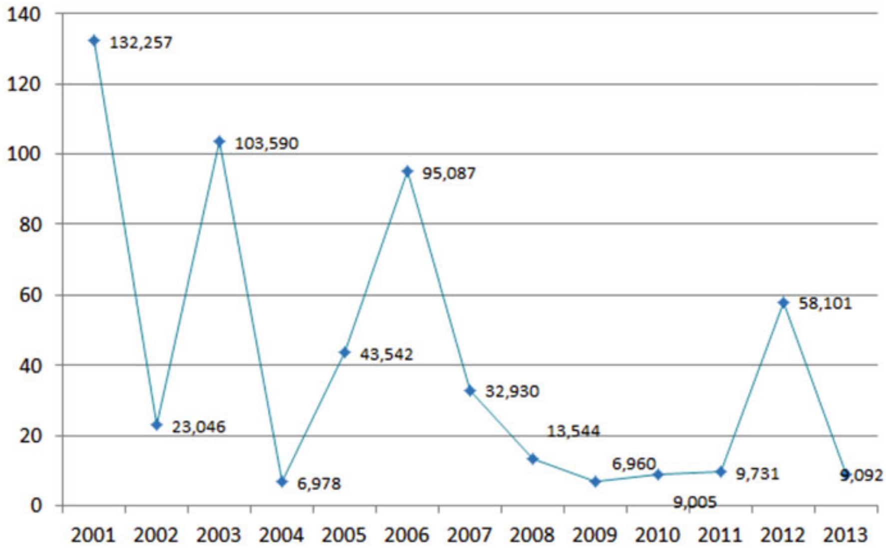


Fig. 5.2 Shapiro–Francia index for departure from normality of German price changes through years

in simple words it is not possible to (linearly) predict future price changes from their past historical values. However, we recall that *seasonality*, *spikes* and *volatility clustering* in electricity prices are empirical facts extensively investigated into the literature.[3] The variance measured by squared price changes displays positive correlation with its own past, hence supporting the time–varying volatility property. Finally, the unconditional distribution of daily price changes does not follow the normal distribution. The Shapiro and Francia (1972) test for normality has been performed given that we have more than 2000 but less than 5000 observations. The corresponding index represented in Fig. 5.2 indicates departure from normality when its values are large; and indeed, all its p–values across years confirm non–normality at 1% confidence level, even if the magnitude of this index is decreasing over time. Descriptive statistics of price levels and differences are presented in Table 5.1. We clearly have an average of 260 observations per year, with price levels, as well as price changes, showing values of the skewness significantly different from zero and values of the kurtosis significantly higher than three.

After having studied the empirical autocorrelation and partial autocorrelation functions, and accounting for the time–structure of the data, we select the autoregressive processes with five lagged price differences for all considered models.

[3]Electricity prices are affected by different forms of seasonality: as the intra–daily one (for night–time and day–time), weekly seasonality reflecting no business activities during weekends; and the more general calendar seasonality, with summers characterized by high electricity demand for air conditioning. Whereas, spikes are abnormal upward and downward price movements.

Table 5.1 Descriptive Statistics for German day–ahead price levels and differences

Year	N	Mean	StdDev	Skewness	Kurtosis	Max	Min
Price levels							
2001	261	26.69	310.82	9.16	101.70	240.26	9.83
2002	261	25.27	64.41	1.21	6.22	61.00	3.47
2003	261	33.42	174.33	4.48	40.99	163.46	3.54
2004	262	31.19	25.84	−0.53	5.37	46.61	12.06
2005	260	50.65	371.58	2.49	11.08	145.97	19.83
2006	260	57.03	661.91	4.80	40.09	301.54	19.79
2007	261	42.78	441.78	2.27	9.68	158.97	5.78
2008	262	71.70	277.92	0.17	3.70	131.40	22.70
2009	261	42.09	112.43	1.00	5.15	86.36	7.21
2010	261	47.50	52.30	0.49	4.44	72.06	21.05
2011	260	54.03	37.60	−0.98	5.27	68.30	27.67
2012	261	46.09	161.23	−3.04	30.25	98.98	−56.87
2013	261	47.85	158.01	0.66	5.12	91.89	14.20
First differences							
2001	260	0.04	15.67	3.52	91.39	182.19	−131.35
2002	260	−0.05	6.58	0.53	8.61	33.07	−24.37
2003	206	0.05	14.81	1.77	60.05	148.18	−124.20
2004	261	−0.02	4.66	−0.07	4.97	15.98	−18.82
2005	259	0.07	11.10	1.42	17.04	73.46	−51.69
2006	259	0.00	24.34	0.08	39.39	200.80	−191.22
2007	260	−0.03	14.68	−1.15	10.86	62.37	−79.15
2008	261	0.05	10.43	0.58	7.93	60.04	−34.43
2009	260	−0.05	6.85	0.12	5.15	28.79	−26.96
2010	260	0.07	4.98	0.42	5.67	19.94	−17.64
2011	259	−0.03	5.23	−0.11	6.60	25.05	−23.48
2012	260	−0.12	10.24	−2.08	36.59	66.21	−95.02
2013	260	0.06	8.01	−0.57	6.37	24.16	−43.04

5.4 Model Setting and Estimation

Even if VaR analysis is based on returns, we were forced to compute simply price differences because electricity price series can exhibit null or even negative values. Therefore, we consider the series of daily electricity spot prices and denote the *price change* between day $t-1$ and day t as

$$X_t = \Delta S_t = S_t - S_{t-1}.$$

At date t, we need to predict a (conditional) VaR for X_{t+1}. More precisely, given $\alpha \in (0, 1)$, we are interested in computing

$$\mathrm{VaR}_\alpha(X_{t+1} \mid \mathcal{I}_t),$$

where $\mathcal{I}_t$ is the information at time t, by means of the *GARCH–type* models which are considered a *standard* approach for VaR estimation, where only the conditional mean and standard deviation of returns are estimated. The original GARCH models assumed normal conditional distributions; recently, these models have been expanded to account for skewness and extra-kurtosis, hence supporting more complex distributions, but the additional estimated parameters are not time-varying, opposite to the dynamics of the conditional mean and variance.

Also, we have decided not to include relations with fundamental drivers or additional factors affecting electricity prices in order to provide an overall quantification of model risk. This figure will naturally include all other forms of risks that influence electricity prices, such as those related to regulatory, structural, and fundamental factors.

Taking advantage of the long history of prices available from 2001 to 2013, we investigate the "time evolution" of the measure of model risk across years by adopting a rolling window procedure, as explained in the following sections.

5.4.1 The GARCH Methodology

Let us remind that we consider a particular set of parametric models within the GARCH framework, using the AR(5)-GARCH(1,1) specification, then for any t we have

$$X_t = \mu_t + \sigma_t Z_t, \tag{5.3}$$

where

$$\mu_t = \overline{\mu} + \sum_{i=1}^{5} \phi_i X_{t-i}, \tag{5.4}$$

is the conditional mean expressed as an AR(5) process, and

$$\sigma_t^2 = \omega + \alpha (X_{t-1} - \mu_{t-1})^2 + \beta \sigma_{t-1}^2 \tag{5.5}$$

is the conditional variance in form of a GARCH(1,1) model. Parameters $\overline{\mu}$, ϕ_i, ω, α and β have to satisfy known constraints. Finally, the innovations Z_t form an IID sequence with standard (i.e. mean 0 and variance 1) common distribution. As a consequence, μ_t and σ_t^2 are the conditional (on $\mathcal{I}_{t-1}$) mean and variance, respectively.

Parameters are estimated by ML, i.e. by maximizing the joint likelihood of the variables (X_t) on a given data window. We stress that the form of the likelihood depends on the distribution F_Z of the innovations. As a consequence, different sets of parameters are retrieved using different hypotheses on F_Z, even if the same data window is used for the estimation process.

A technical remark is in order. The joint distribution of a process X_t, following an AR(5)-GARCH(1,1) model as in (5.3), (5.4), and (5.5), is completely specified only when the common standard distribution of innovations F_Z and the parameters ($\overline{\mu}$, ϕ_i, etc.) are fixed. We can associate a probability measure Q to each complete specification such that the process X_t has the given joint distribution under Q. Moreover, in view of the fact that the estimation process (Maximum Likelihood on a given data window) and the parameters are completely determined once the distribution is fixed, we can safely associate a completely specified AR(5)-GARCH(1,1) model to any possible innovations distribution F_Z; hence, a probability Q on $(\Omega, \mathcal{F})$. As a consequence, in what follows a "model" Q will be given directly in terms of the distribution of Z.

The following alternative distributions for Z are considered in the present study[4]:

- the *normal* (N)
- the *student–t*, *skew student–t*, *skew normal*, (ST, SST, SN);
- the *generalized* and *skew generalized* error distributions, (GED and SGED);
- the *Johnson's SU*, (JSU);
- the *normal inverse gaussian*, (NIG);
- the *generalized hyperbolic* and *generalized hyperbolic student–t*, (GHYP and GHST).

For all the listed distributions, the standard versions (i.e. the ones having mean 0 and variance 1) are considered. Hence, we consider the following finite set of models

$$\mathcal{Q} = \{NO, SN, ST, SST, GED, SGED, JSU, NIG, GHYP, GHST\}$$

containing the reference one, Q_0, to be specified below, and the other Qs associated to the above listed alternative distributions for Z.

We believe that the alternative models considered in this analysis constitute a sufficiently representative range of distributions, often used in practice. In particular, we consider distributions that are different from the normal because they account for asymmetry and/or thick tails, two important features to be considered for proper Risk Management. Moreover, we think that the inclusion of additional *realistic* distributions would not affect much our results, even though this is our conjecture not empirically verified.

Note that the listed distributions may depend on additional parameters: for instance, the student–t distribution depends on the number of degrees of freedom.

[4]For the densities of these distributions please refer to Ghalanos (2014).

These parameters[5] are also estimated through Maximum Likelihood on a common data window. As a consequence, we can think that once the parametric class of distributions (ST, SST, etc.) is considered, additional parameters are implicitly determined. In other words, a "model" Q corresponds to the unique standard distribution in a parametric class (ST, SST, etc.), with additional parameters determined by ML estimation. This "identification" will be tacitly understood in what follows, particularly when presenting the estimation procedure in Sect. 5.4.2.

Making use of the GARCH models and using the basic properties of VaR, for each model we have

$$\mathrm{VaR}_\alpha(X_{t+1} \mid \mathcal{I}_t) = -\left(\mu_{t+1} + \sigma_{t+1} q_\alpha(Z)\right).$$

Notice that the parameters $\overline{\mu}$, ϕ_i, ω, α and β will enter the preceding equation through the quantities μ_{t+1} and σ_{t+1}, while the distribution of Z, completely specified by a parametric class, as explained above, will affect $q_\alpha(Z)$. Summary results of the estimates for the AR(5)-GARCH(1,1) models, using the entire time series of 3391 observations, and under different innovations distributions are reported in Table 5.2. We firstly observe that the unconditional mean, $\bar{\mu}$, is never significant, as expected from the descriptive statistics of price changes. Secondly, the autoregressive structure is an important fact to be included, as well as the heteroskedasticity. On the contrary, estimates for skewness (ν) and kurtosis (τ) turn from significant to non-significant values, according to the distribution considered; and they are definitively not important when the hyperbolic distribution is employed. Turning our attention to the ability of each model to fit the data, we can observe without surprise that the worst fitting models are the one with the *normal* (NO) and the *skew normal* (SN) distributions, for which all information criteria show the highest values. On the contrary, the best fitting model is the one with the *Generalized Error Distribution* (GED), where instead all information criteria exhibit the lowest values. As a robustness check, we estimated two variants of the AR(5)-GARCH(1,1) model, in order to take into account two possible facts: (*i*) electricity price levels may be affected by past price variability, that is, today prices are affected by yesterday price movements reflecting possible problems of "scarcity" in the system (problems referred to unavailable capacity, outages, or excess of demand, among others); and, (*ii*) volatility in electricity prices is generally stronger when prices are high. Therefore, we have verified (*i*) by including the standard deviation, as obtained from the conditional variance equation, in the conditional mean equation hence formulating a *GARCH-in-mean* and estimating the AR(5)-GARCH-M(1,1) models defined by (5.3), and (5.4) and (5.5) replaced by

[5]In general, there will be a parameter that controls for skewness and/or one that controls for kurtosis. We denote them, respectively, ν and τ across different families. The family GHYP has a third parameter, that we denote λ.

Table 5.2 Estimates of the AR(5)-GARCH(1,1) models. Note that ν and τ are the parameters related to skewness and kurtosis, respectively, whereas λ is the third parameter for the GHYP distribution. ***, **, and * mean significant at 1%, 5% and 10% confidence levels respectively. AIC, SBC and HQ represent the Akaike, Bayes and Hannan-Quinn information criteria

	NO		SN		ST		SST		GED		SGED		JSU		NIG		GHYP		GHST	
$\bar{\mu}$	0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000	
ϕ_1	−0.445	***	−0.571	***	−0.392	***	−0.391	***	−0.407	***	−0.434	***	−0.413	***	−0.406	***	−0.416	***	−0.431	***
ϕ_2	−0.343	***	−0.352	**	−0.274	***	−0.271	***	−0.260	***	−0.301	***	−0.324	***	−0.293	***	−0.269	***	−0.312	***
ϕ_3	−0.227	**	−0.261	**	−0.114	***	−0.116	***	−0.184	**	−0.220	***	−0.220	***	−0.196	***	−0.153	***	−0.138	***
ϕ_4	−0.228	**	−0.316	**	−0.111	***	−0.096	***	−0.111		−0.126	**	−0.112	***	−0.121	***	−0.078	**	−0.160	***
ϕ_5	−0.102	*	−0.233	**	−0.075	***	−0.083	***	−0.079		−0.049		−0.076	**	−0.094	***	−0.075	**	−0.074	***
ω	0.662	*	0.162		0.152		0.151		0.186		0.151		0.147		0.150		0.152		0.161	
α	0.128	***	0.084	***	0.063	**	0.062	*	0.100	***	0.063		0.057		0.060	*	0.063	**	0.075	***
β	0.869	***	0.892	***	0.901	***	0.901	***	0.886	***	0.899	***	0.901	**	0.901	***	0.900	***	0.897	***
ν			1.028	***			0.960	***			0.994	***	−0.045		0.005		−0.009		−0.895	*
τ					4.207	***	4.214	***	1.045	***	0.972	***	1.385	**	0.452	***	1.841		8.197	***
λ																	−1.273	***		
AIC	7.0237		7.2328		6.6683		6.6723		6.6353		6.7163		6.6897		6.7191		6.7099		6.7090	
SBC	7.0400		7.2509		6.6864		6.6923		6.6534		6.7362		6.7096		6.7390		6.7316		6.7289	
HQ	7.0295		7.2392		6.6748		6.6795		6.6417		6.7234		6.6968		6.7262		6.7176		6.7162	

$$\mu_t = \overline{\mu} + \sum_{i=1}^{5} \phi_i X_{t-i} + \delta \sigma_{t-1},$$

where δ is an additional parameter, and

$$\sigma_t^2 = \omega + \alpha (X_{t-1} - \mu_{t-1})^2 + \beta \sigma_{t-1}^2$$

Also, we have tested (*ii*) by including the price range of the 24 hourly prices observed on the previous day as an explanatory variable in the conditional variance equation of the GARCH models. The price range is given by $PR_{t-1} = \max_h(S_{t-1}^h) - \min_h(S_{t-1}^h)$, where S_{t-1}^h represents the hourly electricity prices observed during hour $h = 1, \ldots, 24$ and day $t-1$. We therefore considered an AR(5)-GARCHX(1,1) model, defined by (5.3) and (5.4), and (5.5) replaced by

$$\sigma_t^2 = \omega + \alpha (X_{t-1} - \mu_{t-1})^2 + \beta \sigma_{t-1}^2 + \gamma PR_{t-1}$$

where $\gamma \geqslant 0$ is an additional parameter.

The estimated coefficients for investigated facts are reported in Table 5.3, but complete results are available on request. These results clearly show how both facts are not influential and can be neglected in the following implementation of our models.

Table 5.3 ML estimates for the $\hat{\delta}$ and $\hat{\gamma}$ coefficients (with p-values in brackets), when respectively price levels are affected by past price volatility – by means of the AR(5)-GARCH-M(1,1) model – and when price volatility is affected by high past prices – by means of the AR(5)-GARCHX(1,1) model. Note that * indicates that convergence was obtained with less strict criteria of tolerance compared with those used for the other models

Innovations	$\hat{\delta}$	(p-value)	$\hat{\gamma}$	(p-value)
NO	−0.014	(0.286)	0.000	(1.000)
SN	0.003	(0.814)	0.000	(0.999)
ST	0.002	(0.914)	0.000	(1.000)
SST	−0.005	(0.805)	0.000	(1.000)
GED	−0.003	(0.368)	0.000	(1.000)
SGED	−0.001*	(0.281)	0.000	(1.000)
JSU	−0.004	(0.853)	0.000	(0.999)
NIG	0.001	(0.970)	0.000	(0.999)
GHYP	0.003	(0.894)	0.000	(1.000)
GHST	0.003	(0.874)	0.000	(1.000)

5.4.2 Dynamic Model Risk Quantification

Once Q_0 and $\mathcal{Q}$ have been specified (up to additional parameters as explained before), a rolling window procedure is recursively adopted to investigate the time evolution of the relative measure of model risk. In particular, for any model $Q \in \mathcal{Q}$ for the innovations, at day t we estimate the parameters of the GARCH model by ML, using a rolling window of the past 260 days. Hence, we are able to forecast the (conditional) distribution of X_{t+1} and then retrieve the $VaR_{Q,\alpha}$, for both values of α, that is 1% and 5%. Hence, we compute the RMMR measure at any date and for both levels of α. Explicitly, the step-by-step procedure is as follows:

1. start at day $t = 265$;
2. consider the estimation window $\mathcal{W}_t = (X_{t-259}, \ldots, X_t)$; note that when $t = 265$, $\mathcal{W}_{265} = (X_6, \ldots, X_{265})$ as the first 5 observations have to be excluded from the estimation step, due to the AR(5) part of the model;
3. for any model $Q \in \mathcal{Q}$ for the innovations (including the reference one Q_0), estimate the parameters of the AR(5)-GARCH(1,1) model for X_{t+1}, specified as in Eqs. (5.3), (5.4), and (5.5), including additional parameters of the innovation distribution (e.g. the degrees of freedom in the t–student model), using ML on the estimation window $\mathcal{W}_t$. We perform this step using the software R with the "*rugarch*" package[6];
4. having the (conditional) distribution $F_{t+1,Q}$ of X_{t+1} under Q, compute

$$\mathrm{VaR}_{\alpha,t+1,Q} = \mathrm{VaR}_{\alpha,Q}(X_{t+1} \,|\, \mathcal{I}_t) = -\left(\mu_{t+1} + \sigma_{t+1} q_{\alpha,Q}(Z)\right)$$

   for $\alpha = 1\%$ and $\alpha = 5\%$, and then

$$\mathrm{RMMR}_{\alpha,t+1} = \frac{\overline{\mathrm{VaR}}_{\alpha,t+1} - \mathrm{VaR}_{\alpha,t+1,Q_0}}{\overline{\mathrm{VaR}}_{\alpha,t+1} - \underline{\mathrm{VaR}}_{\alpha,t+1}}$$

   where

$$\overline{\mathrm{VaR}}_{\alpha,t+1} = \max_{\mathcal{Q}} \mathrm{VaR}_{\alpha,t+1,Q}, \qquad \underline{\mathrm{VaR}}_{\alpha,t+1} = \min_{\mathcal{Q}} \mathrm{VaR}_{\alpha,t+1,Q};$$

5. increment t by 1 day and go back to step 2: in particular, we have $t = 266$ and $\mathcal{W}_{266} = (X_7, \ldots, X_{266})$ at the second iteration; and so on.

Recalling that the entire time series at our disposal has 3391 ($= 3392 - 1$) observations for price changes, the procedure outlined above yields two series of 3126 measures of model risk (with respect to Q_0 and $\mathcal{Q}$):

$$\mathrm{RMMR}_{\alpha,266}, \ \ldots, \ \mathrm{RMMR}_{\alpha,3391} \qquad \alpha = 1\%, 5\%$$

[6] See Ghalanos (2014) for details.

Notice that the actual distribution associated to Q_0 (or any other $Q \in \mathcal{Q}$) will change from one day to another, due to different estimates of the GARCH parameters and additional parameters of innovations. As a consequence, any series $(\mathrm{RMMR}_{\alpha,t})_{t=266,\dots,3391}$ will in fact describe the relative model risk of sticking, day by day, to a given parametric family for innovations, instead of choosing one of the 9 alternative families.

5.5 Empirical Results

We have implemented the procedure just described with three different choices of the reference model Q_0, and precisely

- Normal (NO)
- Generalized Error Distribution (GED)
- Skew-t distribution (ST)

The Normal assumption for innovations, even though still employed in many "everyday" Risk Management procedures, is a poor description of time-series dynamics, particularly for energy variables.[7] Nevertheless, we considered it as a rough benchmark. The GED distribution is the one that *best fits* our data, while the ST is a common choice in the energy literature.

The two series of RMMR (for $\alpha = 1\%$ and 5%), using the normal as the reference model, are reported in the first row of Fig. 5.3. These two graphs can give only a very rough idea of the dynamics of model risk, because of the high density of points. However, it seems that model at 1% (right graph) roughly moved from *low* to *high model risk*. A similar behaviour is shown by model risk at 1% using the other two reference distributions (we omit the graphs, but see Fig. 5.4).

To get clearer dynamics, we have computed moving averages of daily RMMRs on weekly (based on 5 days), monthly (based on 20 days) and yearly (based on 260 days) basis. The dynamics of these aggregations are reported in the remaining rows of Fig. 5.3. They show the substantial difference between considering $\mathrm{VaR}_{5\%}$ or $\mathrm{VaR}_{1\%}$, with the model risk being substantially higher in the latter case. This comes at no surprise, as the normal distribution has tails which are lighter than most of the alternatives. Secondly, and more importantly, results confirm our intuition that the increasing complexity of market mechanisms and regulation, together with the introduction of new fundamental drivers (wind and solar), have affected the stochastic nature of electricity prices, hence inducing more *model risk* given that the data are becoming progressively less normal through years.

[7]However, we point out that in GARCH models with normal innovations, the variables X_t themselves are not normal, as they always display extra-kurtosis.

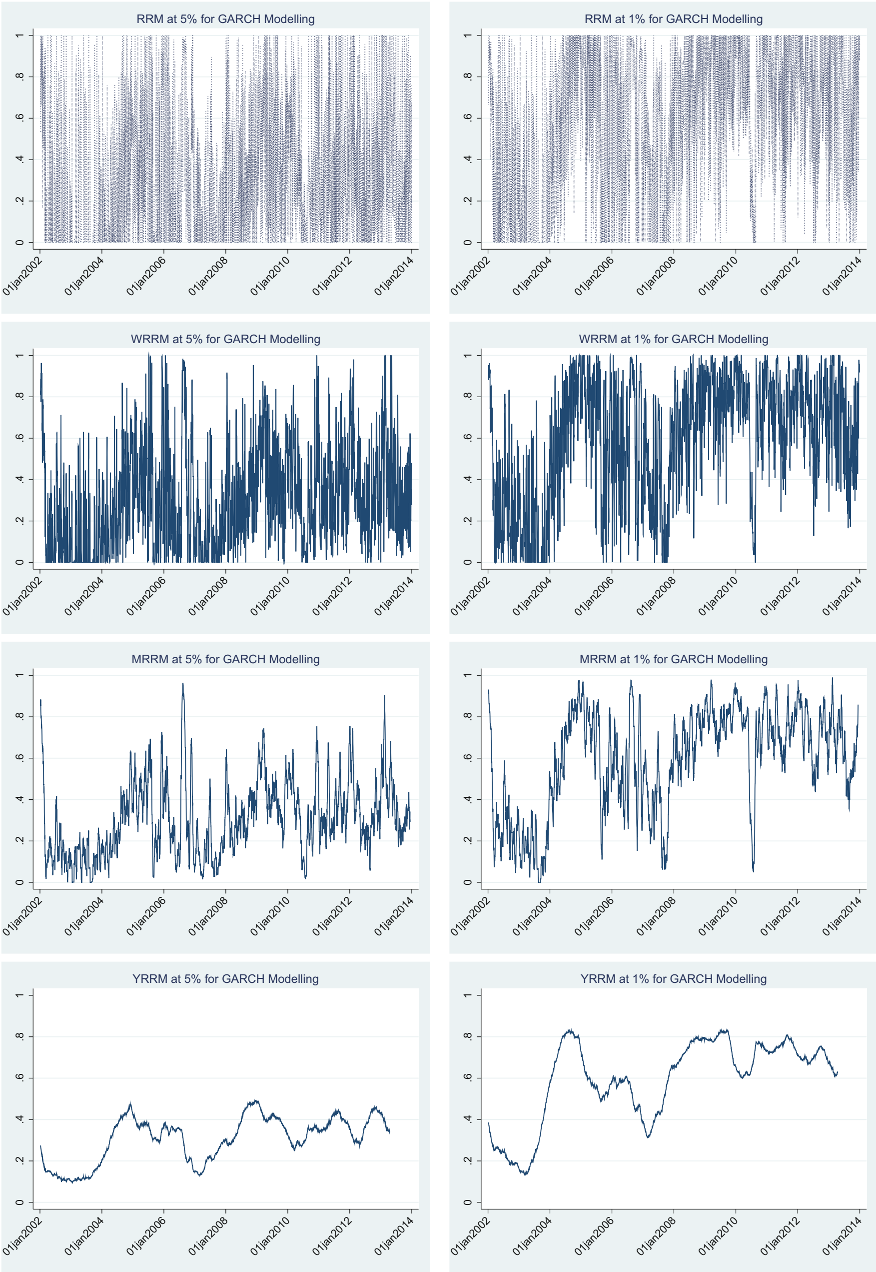
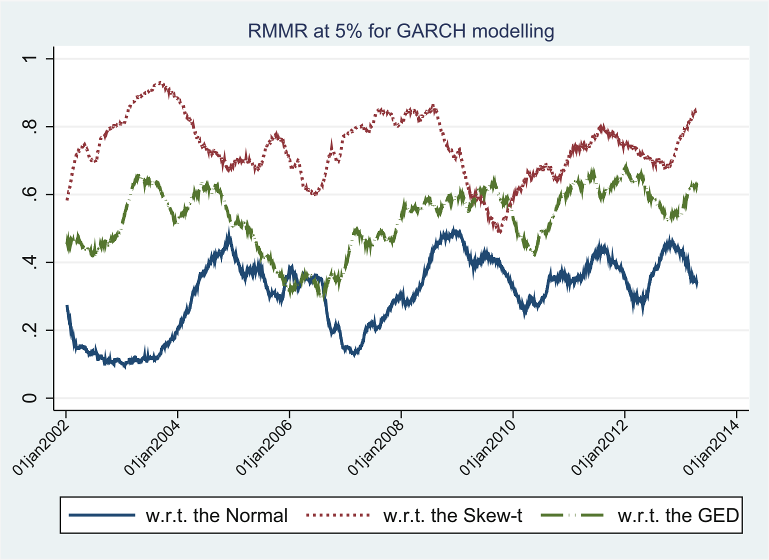


Fig. 5.3 Dynamics of the RMMR series with $\alpha = 5\%$ (*left*) and $\alpha = 1\%$ (*right*) using the GARCH setting with normal (NO) as reference distribution. From *top*: daily values, weekly averages, monthly averages, yearly averages



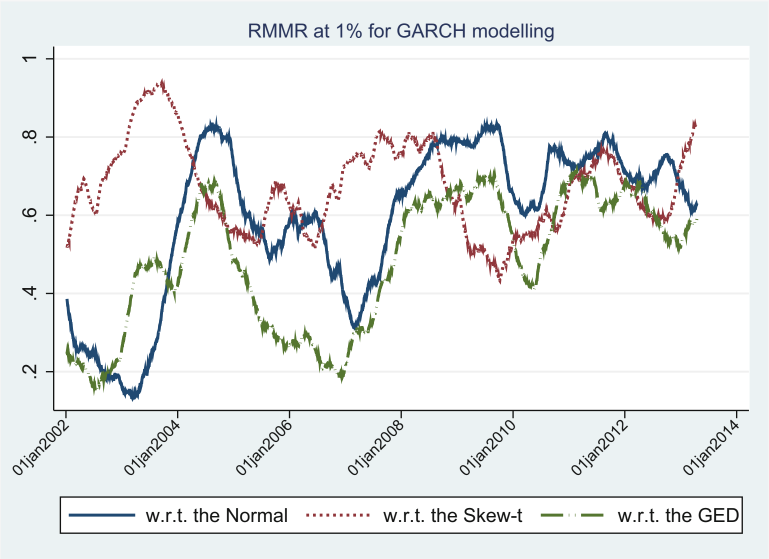


Fig. 5.4 Dynamics of the yearly moving average of the RMMR series with $\alpha = 5\%$ (*top*) and $\alpha = 1\%$ (*down*) using the GARCH setting under three different reference distributions (NO, ST, GED)

A comparison of the RMMR yearly moving averages for $\alpha = 5\%$ and 1% for the three reference models (GED, ST and NO) is reported in Fig. 5.4. We can see that the same rough increase of model risk at 1% is observed for the GED case.

Table 5.4 Fraction of days in which the highest ($\overline{\mathrm{VaR}}_\alpha$) or lowest ($\underline{\mathrm{VaR}}_\alpha$) risk figures are attained by a specific distribution in $\mathcal{Q}$ together with the average values of VaRs ($\widehat{\mathrm{VaR}_\alpha}$) and the number of re-estimated VaRs ($\sharp Obs$)

	$\sharp obs$	$\widehat{\mathrm{VaR}_{5\%}}$	$\overline{\mathrm{VaR}}_{5\%}$ (%)	$\underline{\mathrm{VaR}}_{5\%}$ (%)	$\sharp obs$	$\widehat{\mathrm{VaR}_{1\%}}$	$\overline{\mathrm{VaR}}_{1\%}$ (%)	$\underline{\mathrm{VaR}}_{1\%}$ (%)
NO	3126	12.872	33.08	4.73	3126	18.226	13.95	16.31
ST	3126	9.801	1.18	18.68	3126	17.170	1.82	14.91
SST	3126	9.987	1.95	10.84	3126	17.554	3.23	6.75
SN	3126	12.191	20.51	5.34	3126	17.025	3.84	18.91
GED	3126	11.384	8.61	7.26	3126	19.668	15.23	7.07
SGED	3126	10.955	6.37	6.59	3126	18.869	7.10	4.25
JSU	3126	10.315	3.68	18.04	3126	18.661	8.77	4.38
GHST	3126	11.310	13.50	3.23	3126	19.074	16.67	1.44
GHYP	3126	10.237	2.78	9.47	3126	16.990	5.60	17.85
NIG	3126	10.348	8.35	15.80	3126	18.901	23.80	8.13
			100.00	100.00			100.00	100.00

It is worth stressing that an high fitting ability of the reference distribution does not necessarily mean a low model risk. This is clear from Fig. 5.4 where the $\mathrm{RMMR}_{5\%}$ for the GED reference model is significantly higher than the one for the normal reference model, despite the superior ability of the GED distribution in fitting the considered data. Indeed, a poor fitting can yield both to an underestimation or an overestimation of VaR: in the former case, the model risk is increased, whereas in the latter case the model risk is decreased. For instance, it is a known fact that, in general, VaR_α of an equity portfolio is significantly overestimated at $\alpha \gtrsim 5\%$ and underestimated for $\alpha \lesssim 5\%$, if a normal distribution is used for returns.

Finally, we have also computed at how many days a given model Q attains the highest or the lowest VaR, i.e. $\mathrm{VaR}_{t,Q} = \overline{\mathrm{VaR}}_t$ or $\underline{\mathrm{VaR}}_t$. For any model, the related frequencies are presented in Table 5.4 together with the average values of VaRs. These results clearly show that none of the alternative models attains the maximum or minimum VaRs too frequently: this seems to suggest that the final model risk figures are not consistently biased by any of the selected models.

5.6 Conclusions and Future Research

We provide for the first time an empirical assessment of model risk in the German wholesale electricity market implementing a quantitative measure of model risk for VaR, proposed by Barrieu and Scandolo (2015). The German electricity market has undergone deep structural changes as the modification of the generation mix required to reduce carbon emissions, and this has increased the complexity of the energy sector and affected the stochastic nature of electricity prices, which are characterized by new empirical distributional shapes. Therefore, relaxing the

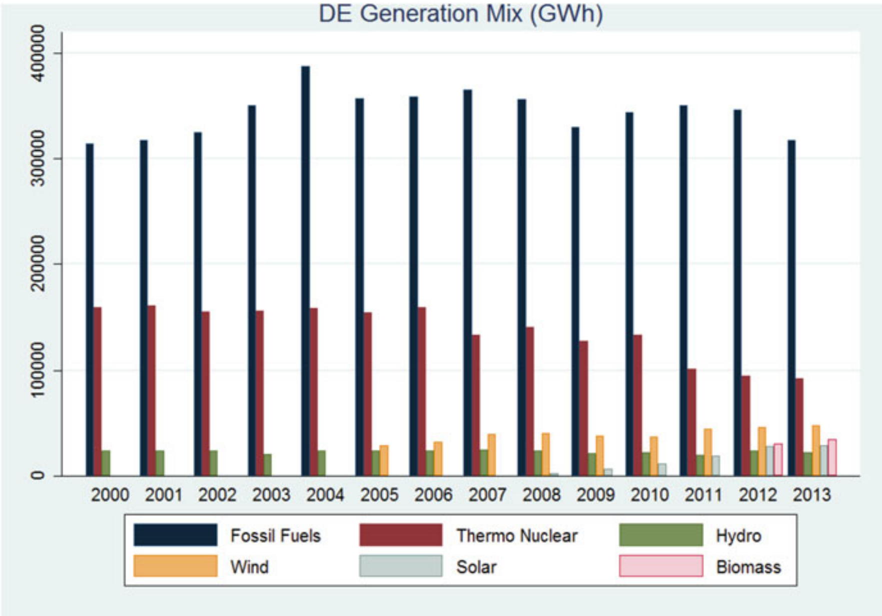


Fig. 5.5 German generation mix over years

assumption of normality and using a wide range of alternative distributions, we have quantified model risk under the well-established setting of GARCH models.

These results emphasize that the distributional assumptions made in price modelling can produce a relevant difference and then trigger substantial model risk. In this specific case, rolling measures, computed with respect to the normal distribution, exhibit average values very close to one for more recent years; hence, indicating *full model risk*.

Indeed, it is interesting to observe that there has been a progressive increase in *model risk* from 2007. This may be explained looking at the evolution of the German generation mix, and analyzing the data collected from ENTSO–E.[8] Figures 5.5 and 5.6 clearly show the important transformations occurred in this power market: starting from 2007 there has been a progressive reduction of nuclear generation, and a pretty constant generation from hydro, but an increasing share of wind together with a substantial amount of electricity produced by solar. Furthermore, it is important to recall that the negative pricing has been introduced on the intra–day market in 2007 and on the day–ahead German/Austrian market in 2008; and this has substantially modified the empirical properties of hourly electricity prices, and consequently daily average prices.

[8]ENTSO-E is the European network of transmission system operators for electricity. More information can be found at www.entsoe.eu.

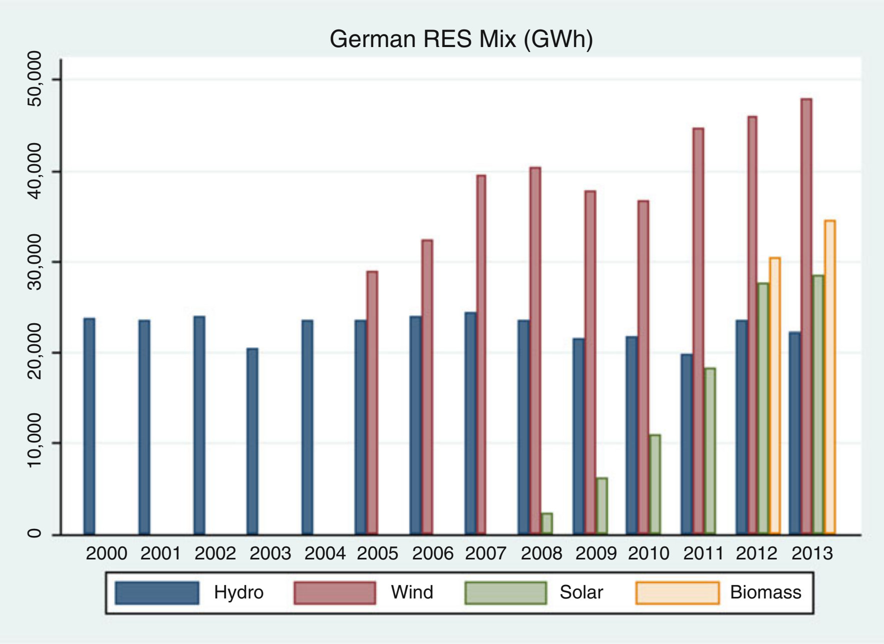


Fig. 5.6 German RES mix over years

Although our results can be considered an exploration, they clearly show that the increasing complexity of market mechanisms and their regulation, together with the introduction of new fundamental drivers as renewable energy sources (specifically wind and solar), have dramatically modified the nature and statistical properties of electricity prices. Indeed, we document a progressive increase of model risk with the time–evolution of the generation mix, and under three different assumptions for the reference model. This fact highlights the importance of considering the impact that the model choice has on risk assessment, particularly when operating in complex markets such as the energy ones.

To this aim, this analysis can be considered "preliminary" and future developments will address at least two issues. First, we plan to extract a "net model risk" by explicitly modelling, hence excluding, other sources of risks such as the *demand*, *RES-induced*, and *fundamental* risks related to fuels. Second, we aim at implementing a more flexible methodology which takes into account time-varying moments depending on plausible energy risk factors. We have explored the great potential of a recent methodology (the *generalized additive models for location, scale and shape*, GAMLSS[9]), encompassing equations for location, scale and shape. Indeed, we have considered the simplest formulation with constant variance,

[9]See Rigby and Stasinopoulos (2005) for a full description with all details.

skewness and kurtosis in a preliminary analysis, and results on RMMR obtained within this framework recall the dynamics of those obtained with the GARCH models, hence providing the same qualitative indications even if with obvious different magnitudes due to differences in model formulations.[10] Clearly, we expect to get a better picture of the model risk once the other sources of electricity risk are explicitly included in this more flexible framework, and once an autoregressive structure is adopted in the equations of all four moments.

To conclude, we have emphasized that the risk quantification of a financial position crucially depends on the employed model. If a set of alternative models is fixed, including all possible distributions that is meaningful to consider, then the measure of model risk we have considered can give useful indications about the model which is currently implemented (i.e. the *reference* one). If RMMR has been consistently close to 1 in the recent period, then it is likely that VaR will be underestimated in the subsequent days, if we use the reference model. In this case, proper actions can be taken, such as incrementing the VaR figure by an amount that depends on model risk. In practice, implementing sophisticated models can be costly, for instance, in terms of IT equipments, estimation times, and need of skilled employees. In view of these considerations, the measure of model risk could be computed for a limited amount of time and from time to time. Still, this would provide risk managers with a useful figure that could be complementary to the usual back-testing results for VaRs.

Acknowledgements We thank two anonymous referees who helped us to improve the paper. Furthermore, the first author acknowledges the European financial support under the scheme *Marie Curie Actions*, *Intra–European Fellowship for Career Development* (FP7–PEOPLE–2011–IEF) for *AaDAMS*.

References

K. Bannör, R. Kiesel, A. Nazarova, M. Scherer, Parametric model risk and power plant valuation. Energy Econ. **59**, 423–434 (2016)

P. Barrieu, G. Scandolo, Assessing financial model risk. Eur. J. Oper. Res. **242**(2), 546–556 (2015)

M. Bhattacharyya, G. Ritolia, Conditional VaR using EVT towards a planned margin scheme. Int. Rev. Financ. Anal. **17**, 382–395 (2008)

Y. Fan, Y.J. Zhang, H.T. Tsai, Y.M. Wei, Estimating 'Value at Risk' of crude oil price and its spillover effect using the GED–GARCH approach. Energy Econ. **30**, 3156–3171 (2008)

A. Ghalanos, Introduction to the rugarch package (Version 1.3-1). Technical report v. Available at http://cran.r-project.org/web/packages/rugarch

[10] These results rely on a structure which is simpler than the one presented in this paper, because of the constant higher moments opposite to the time-varying volatility of the GARCH framework. Among other differences, we refer for instance to the linearities between the equations of sample moments, and the presence/absence of their autoregressive structures. These results are available on request.

A. Gianfreda, D. Bunn, Identifying the dynamic shape–shifters in electricity price densities. EMG working paper (2015)

REN21, Renewable Energy Policy Network for the 21st century, 2014. Renewables 2014 – global status report. Available at http://www.ren21.net/

R.A. Rigby, D.M. Stasinopoulos, Generalised additive models for location, scale and shape (with discussion). Appl. Stat. **54** 1–38, (2005)

S.S. Shapiro, R.S. Francia, An approximate analysis of variance test for normality. J. Am. Stat. Assoc. **67**, 215–216 (1972)

Part II
Pricing and Valuation

Chapter 6
Wine Futures: Pricing and Allocation as Levers Against Quality Uncertainty

Tim Noparumpa, Burak Kazaz, and Scott Webster

Abstract This study examines the impact of using wine futures in order to mitigate the winemaker's risk stemming from quality uncertainty. In each vintage, a winemaker harvests grapes and crushes them in order to make wine. A premium wine sits in barrels for 18–24 months. During the aging process, tasting experts take samples and establish a *barrel score*; this barrel score often indicates the expert's perception of whether the wine will be a superior wine. Based on the barrel score, the winemaker can sell some or all of her/his wine in the form of wine futures and in advance of bottling. The winemaker makes three decisions: (1) the price to sell her/his wine futures, (2) the quantity of wine futures to be sold in advance, and (3) the amount of wine to be kept for retail and distribution. The wine continues to age for one more year after barrel samples. The tasting experts then provide a *bottle score* upon the bottling of the wine. At the time the winemaker determines the price and quantity of wine futures, this unrealized bottle score represents the uncertainty that influences the market price of the wine.

This study makes two contributions to the optimization of pricing and quantity decisions and offers insightful recommendations for practicing managers. First, it develops a stochastic optimization model that integrates uncertain consumer valuations of wine both in the form of futures and in bottle, and the uncertainty associated with bottle scores. Second, it provides an empirical analysis using data collected from Bordeaux wineries engaging in wine futures. The empirical analysis demonstrates that wine futures can be used as price and quantity levers to mitigate the negative consequences of quality uncertainty. The results provide clues as to how

T. Noparumpa (✉)
Chulalongkorn Business School, Chulalongkorn University, Bangkok, Thailand 10330
e-mail: tim@cbs.chula.ac.th

B. Kazaz
Whitman School of Management, Syracuse University, Syracuse, NY, 13244, USA
e-mail: bkazaz@syr.edu

S. Webster
W.P. Carey School of Business, Arizona State University, Tempe, AZ, 85287, USA
e-mail: scott.webster@asu.edu

G. Consigli et al. (eds.), *Handbook of Recent Advances in Commodity and Financial Modeling*, International Series in Operations Research & Management Science 257,
DOI 10.1007/978-3-319-61320-8_6

other markets (e.g. Italy and the U.S.) can establish similar wine futures markets in order to help their small and artisanal winemakers.

Keywords Stochastic optimization • Wine futures • Futures pricing • Quality uncertainty

6.1 Introduction

This study shows how wine futures can be used as pricing and quantity levers in order to mitigate the negative consequences of quality uncertainty in the process of winemaking. Selling wine in the form of wine futures refers to the winemaker's decision to offer her/his wine in advance and prior to the completion of the winemaking process. Fine wine often requires a long aging process for the liquid resting in barrels, e.g. 18–24 months for most Bordeaux style wines. Thus, a winemaker has her cash tied up in this inventory for a long period of time before the wine gets bottled and distributed for retail. Wine futures can offer the winemaker the opportunity to collect some of this cash investment earlier and transfer a proportion of the risk stemming from uncertain bottle reviews to consumers.

The paper builds a stochastic optimization model in order to assist a winemaker in her decisions regarding the proportion of wine to be sold in advance in the form of wine futures and the price of wine futures. Consequently, the model determines the proportion of the wine that should be distributed for retail sale in later periods. The uncertainty in the life of a fine wine producer arises from the critical reviews of tasting experts. These experts review the wine while it is still aging in barrels and provide a score that indicates projections regarding the quality of the wine. Potential buyers rely on this information in order to determine whether to purchase wine futures. The same tasting experts provide a second review when the winemaking process is completed and the wine is bottled. This bottled wine review can differ from the review provided in the barrel phase. The model in this paper model helps the winemaker to mitigate the negative consequences of the uncertainty stemming from as-yet unknown bottle reviews.

We begin the discussion with the description of the winemaking process.

6.1.1 *Winemaking Process and the Tasting Reviews*

A fine wine producer in the US and in Europe harvests grapes in September and/or October. After crushing grapes and shuffling the juice in a tank (wood, steel, or concrete), red wines are then transferred to oak barrels; this marks the beginning of the aging process. The wine continues its aging process in barrels for 18–24 months.

Tasting experts visit these fine wine producers 6–8 months after harvest (in March and April). These tasting experts provide their reviews and assign a *barrel score*, often out of 100 points. The most influential and widely-distributed magazine

Wine Spectator, for example, describes its 100-point scoring system as follows: A classic (great) wine receives a score between 95 and 100, and an outstanding wine (a wine of superior character and style) 90–94, a very good wine (a wine with special qualities) 85–89 points, a good wine (well-made wine) 80–84 points, a mediocre wine (a drinkable wine that may have minor flaws) 75–79. A wine that receives a score 74 or below is not recommended by *Wine Spectator*.

The same tasting expert provides another score, called *bottle score*, when the wine completes its aging process and gets bottled. This bottle score can differ from barrel score, and is the primary source of the risk for a fine wine producer. This work develops a stochastic optimization model that uses the barrel scores in order to mitigate the uncertainty in bottle reviews.

It is known that premium French wines have been sold in the form of wine futures since the seventeenth century. “En primeur” is the French concept of selling wine while it is still aging in the barrel. En primeur is translated into English as “wine futures” indicating financial contracts with standardized terms. These wine futures are traded in an electronic exchange market called the London International Vintner’s Exchange, or shortly known as “Liv-ex.” This electronic platform is similar to NASDAQ, however, only highly sought-after fine wines are traded both in the form of wine futures and in bottle in Liv-ex. Figure 6.1 demonstrates a screenshot of the trading platform in Liv-ex. Merchants, brokers, retailers, and consumers make up the buyers of wine futures in Liv-ex.

Tasting reviews have significant impact on the quality perception of wine. *Wine Spectator* is the most widely distributed magazine in the wine industry and has a significant impact on the quality perception of wine. Masset et al. (2015) demonstrates that a 10% increase in barrel scores provided by *Wine Spectator* leads to a 4% increase in futures price.

The winemaker and wine futures consumers exhibit distinct properties that differ from the common description of risk aversion in (the industrial organization theory of) economics literature. According to the industrial organization theory,

Fig. 6.1 A screenshot of Liv-ex trading platform

(large) firms can diversify their risk and do not need to behave in a risk-averse manner. The same theory indicates that individual consumers would exhibit a risk-averse behavior as they possess limited resources and cash. However, consumers in the wine industry are affluent collectors and/or financially-healthy merchants and distributors. The empirical analysis in Noparumpa et al. (2015a) does not support a risk-averse behavior on the part of wine futures consumers. Therefore, the consumers of wine futures in this study are considered to be risk-neutral in order to reflect the true operating environment in this industry. Because winemakers are often small in size with limited financial resources, they exhibit a risk-averse behavior. These unique features are incorporated into the empirical analysis as well.

This paper develops a stochastic optimization model that maximizes the percentage profit improvement from using wine futures as price and quantity levers in the presence of bottle score uncertainty. It demonstrates that these two decisions, futures quantity and futures price, serve as effective levers against quality uncertainty. Section 6.2 reviews the corresponding literature. Section 6.3 introduces the model; Section 6.4 presents its analysis. Section 6.5 provides an empirical analysis using data from Bordeaux wineries as well as one of the US artisanal winemakers. Section 6.6 presents the conclusions.

6.2 Literature Review

There are three streams of literature related with this study.

6.2.1 Pricing and Quantity Decisions Under Uncertainty

Supply chain and operations management literature focuses primarily on pricing and quantity decisions under uncertainty. Specifically, the *Price-Setting Newsvendor Problem* (PSNP) investigates the problem of determining a selling price and a production quantity (or inventory level) under demand uncertainty: Van Mieghem and Dada (1999), Petruzzi and Dada (1999), Dana and Petruzzi (2001), Federgruen and Heching (1999, 2002) and Kocabıyıkoğlu and Popescu (2011) are examples of studies that examined these two critical decisions under stochastic demand. These studies, however, assume that supply and quality are deterministic, and thus, supply and quality fluctuations do not influence pricing and quantity decisions. This work differs from these earlier publications in three ways: (1) It incorporates quality uncertainty into the joint pricing and quantity decisions under uncertainty; (2) it develops two new levers with advance selling quantity (i.e., the amount of futures to be sold) and the advance selling price (futures price) as levers to mitigate quality uncertainty; and, (3) it examines a risk-averse winemaker (firm) who is interested in transferring a proportion of her quality risk to consumers. Kazaz and Webster (2015) incorporate supply uncertainty into PSNP, and develop a new elasticity measure leading to unique optimal solutions in the problem of determining price and quantity under supply and demand uncertainty; however, their study does not feature a retail market analysis. Finally, this study shows that advance pricing through futures prices and advance allocation are financially viable risk mitigation techniques for winemakers with a significant amount of cash tied up in inventory that may diminish in value.

In addition to the PSNP literature, there is a growing body of literature that examines the impact of supply uncertainty. Jones et al. (2001), Kazaz (2004), and Kazaz and Webster (2011), Noparumpa et al. (2015b) demonstrate the benefits of using a secondary source of supply in order to mitigate the negative consequences of supply uncertainty. Rather than utilizing a secondary source, this work focuses on the use of advance selling as a lever against uncertainty. Moreover, we examine quality uncertainty rather than supply uncertainty. In this problem, the quality of the final product can fluctuate during the course of the aging process.

6.2.2 Advance Selling

Marketing literature demonstrates the benefits of advance selling as consumers get the opportunity to purchase goods or services before the time of consumption. Gale and Holmes (1992, 1993), Shugan and Xie (2000, 2005), Xie and Shugan (2001), Fay and Xie (2010), Boyaci and Özer (2010), Tang and Lim (2013), Cho and Tang (2013) show that advance selling is a method where a firm can discriminate its consumer base through differential pricing. These studies demonstrate that advance selling helps a firm to manage fluctuations in demand. Hekimoğlu et al. (2017) utilizes advance selling in the context of purchasing wine in advance. Their study examines how a wine distributor can benefit by purchasing wine in the form of futures in addition to bottled wine under a limited budget while complying with the firm's degree of risk aversion. In another study that employs the concept of advance selling, Gheibi et al. (2017) investigate optimal farm leasing and purchasing decisions on behalf of coffee processors. Our study differs from these studies by introducing quality uncertainty at the time of consumption. This work differs from these studies by introducing quality uncertainty at the time of consumption.

6.2.3 Wine Tasting

A wide majority of the economics literature on wine pricing has focused on the influence of weather fluctuations in growing seasons. Ashenfelter (2010) and Ashenfelter et al. (1995), for example, focus on the impact of climactic conditions on the quality and price of aged wines. There is a growing literature examining various aspects of the influence in wine tasting. Ali et al. (2008), Ashenfelter and Jones (2013), Stuen et al. (2015), Bodington (2015), and Olkin et al. (2015) are examples of studies that investigate the impact of wine tasting experts in creating the perception of quality. Masset et al. (2015) examine the influence of various tasting experts on futures prices. However, these publications do not develop a stochastic optimization model for the winemaker in order to determine quantity (the amount of wine to be sold in the form of futures) or price (price of wine futures); thus, they do not emphasize building levers to mitigate quality uncertainty.

6.2.4 Contribution over Noparumpa et al. (2015a)

This paper is similar to Noparumpa et al. (2015a) where they also build an analytical model based on tasting expert reviews. It departs from the earlier publication in three dimensions. First, this work removes the assumption of risk-averse consumers, and considers risk-neutral buyers of wine futures. The considerations of risk-neutral consumers is reflective of the operating environment as the ultimate buyers of futures, as explained before, constitute an affluent customer base. Thus, this study leads to more accurate estimations of the benefits that can be obtained by optimizing the winemaker's pricing and quantity decisions at the futures stage. Second, the emphasis is on the research question associated with identifying and measuring the financial benefit from using a stochastic optimization model. Noparumpa et al. (2015a) do not even report on the benefit of their analytical model. Third, this study employs barrel and bottle scores from *Wine Spectator* and Noparumpa et al. (2015a, b) relies on Robert Parker's reviews.

In sum, this study integrates marketing, economics and supply chain management by studying the price and quantity decisions in the form of wine futures. From a marketing perspective, we show that a futures price can act as a lever to discriminate buyers through futures and retail prices. From an economics perspective, the pricing decision helps the winemaker to extract additional surplus from consumers. From a supply chain management perspective, selling wine in advance of bottling enables the winemaker to transfer the risk of holding inventory that fluctuates in value due to quality-rating uncertainty to buyers of wine futures; it also helps the winemaker to recover some of her cash investment.

6.3 The Model

This section presents the modeling approach useful to a facing quality-rating uncertainty and seeking an optimal supply policy in the spot and futures market. Both the futures price and the wine supply in the futures market are endogenous. We develop a two-stage stochastic programming model in order to formulate the problem for the winemaker. Figure 6.2 depicts the timeline of events and decisions. At the end of grape harvest (September of calendar year t), the winemaker obtains a certain amount of wine described by Q of wine from vintage t. After the wine

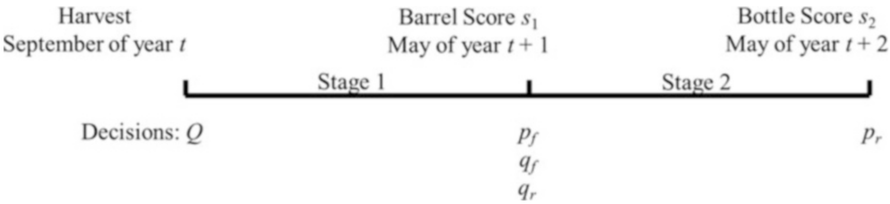


Fig. 6.2 Model specifications, stages in the model, and the timeline of decisions

ages for 8–10 months, tasting experts (e.g. Robert Parker Jr. of *The Wine Advocate*, James Molesworth of *Wine Spectator*, Jancis Robinson of *Financial Times*) visit the winery in order to taste the wine. The tasting expert provides a barrel score in May of calendar year $t + 1$ that gets revealed to the winemaker and consumers through a publication. We denote the realized barrel score assigned by the tasting expert with s_1. The realization of the barrel score marks the beginning of stage 1 in the model. The winemaker then makes the following three decisions based on the realized barrel score:

1. The price of wine futures, denoted p_f, which determines the demand for wine futures, denoted $d_f(p_f)$,
2. The quantity of wine to be sold as futures, denoted q_f,
3. The quantity of wine that is reserved for retail distribution, denoted q_r.

In stage 1, the winemaker sells a quantity of wine in the form of futures equivalent to $\min\{d_f(p_f), q_f\}$ at the unit price of p_f. Thus, the winemaker collects a revenue of $p_f \min\{d_f(p_f), q_f\}$ in stage 1. The remaining portion of wine is distributed for retail sale in the second-stage of the model; we denote this quantity $q_r = Q - \min\{d_f(p_f), q_f\}$. Although the quantity decisions might be in integer values in reality, the model utilizes continuous values, and can be perceived as an approximation to the ideal amount of wine that should be sold in the form of wine futures.

At the end of the aging process, in May of calendar year $t + 2$, the wine gets bottled and goes through another review of the wine-tasting experts. The random bottle score is expressed as $\tilde{s}_2$, and its realization as s_2. Because the barrel score provides an indication of the final bottle score s_2, the random variable $\tilde{s}_2$follows a conditional probability density function $f(s_2 \mid s_1)$. We assume that the expectation of the bottle score in May of calendar year $t + 2$ when the barrel score is revealed in May of calendar year $t + 1$ is equal to the barrel score, i.e., $E\left[\tilde{s}_2|s_1\right] = s_1$.

At the end of Stage 2 of the model, the wine is sold at a retail price p_r that is influenced by the random bottle score that gets revealed in May of calendar year $t + 2$ as well as the barrel score that is revealed in May of calendar year $t + 1$. Without loss of generality, we normalize the bottle price of retail wine to be equivalent to the bottle rating of the wine; specifically, we have $p_r = p_r(s_2) = s_2$. It follows from $E\left[\tilde{s}_2|s_1\right] = s_1$ that the expected retail price in May of calendar year $t + 1$ is equal to the barrel score: $E\left[p_r\left(\tilde{s}_2|s_1\right)\right] = s_1$. As a consequence, the winemaker collects a revenue equivalent to $E\left[p_r\left(\tilde{s}_2|s_1\right)\right]\left(Q - \min\left\{d_f\left(p_f\right), q_f\right\}\right)$ in stage 2.

The revenues collected in stage 2 are discounted to May of calendar year $t + 1$ through the winemaker's attitude towards risk under uncertainty. We adopt the risk-adjusted discount rate that is common in the finance literature (e.g., Samuelson 1963). The value of the risk-adjusted discount rate, denoted ϕ, depends on the risk of selling a bottle of wine at an uncertain retail price in the future. The higher the uncertainty in bottle price and the more risk-averse the winemaker is, the lower the value of ϕ. If the winemaker is risk-neutral, for example, then $\phi = (1 + r)^{-1}$. The winemaker's risk-adjusted expected profit for a given set of first-stage decisions (p_f, q_f) can then be expressed as follows:

$$\Pi\left(p_f, q_f\right) = p_f \min\left\{d_f\left(p_f\right), q_f\right\} + \phi E\left[p_r\left(\tilde{s}_2|s_1\right)\right]\left(Q - \min\left\{d_f\left(p_f\right), q_f\right\}\right). \quad (6.1)$$

6.3.1 The Model

We develop a model that maximizes the benefits from using the optimal choices of futures price and quantity of wine to be sold in the form of wine futures. We describe the winemaker's present choice of futures price and quantity with p_f^0 and q_f^0, respectively. For the futures price p_f^0, the demand for wine futures is $d_f\left(p_f^0\right)$. Thus, the present profit level for the winemaker can be expressed as:

$$\Pi\left(p_f^0, q_f^0\right) = p_f^0 \min\left\{d_f\left(p_f^0\right), q_f^0\right\} + \phi E\left[p_r\left(\tilde{s}_2|s_1\right)\right]\left(Q - \min\left\{d_f\left(p_f^0\right), q_f^0\right\}\right).$$

This model maximizes the benefits exceeding the present level of profitability. We describe the percentage of profit improvement through the proposed modeling approach as:

$$\Delta\Pi\left(p_f, q_f\right) = \left(\Pi\left(p_f, q_f\right) - \Pi\left(p_f^0, q_f^0\right)\right) / \Pi\left(p_f^0, q_f^0\right). \quad (6.2)$$

We can now express the model as follows:

$$\max_{\left(p_f, q_f\right) \geq 0} \Delta\Pi\left(p_f, q_f\right) \quad (6.3)$$

s.t.

$$q_f \leq Q. \quad (6.4)$$

It is important to observe that the winemaker's current choice of (p_f^0, q_f^0) can differ from the optimal decisions in the price and quantity pair (p_f, q_f). Specifically, the winemaker might have chosen a futures price to be higher (or lower) than the optimal futures price, i.e., $p_f^0 > p_f$ (or $p_f^0 < p_f$) leading to a lower (or higher) wine futures demand than ideal. Similarly, the winemaker's current allotment as futures q_f^0 can be higher or lower than the ideal quantity of futures q_f. Thus, the model in

(6.3) – (6.4) leads to the highest benefit from a simultaneous optimization of price and quantity decisions for the futures market. In reality, the firm might determine integer values for its quantity decisions, however, the model uses continuous decision variables. Therefore, the results should be perceived as an approximation of the benefits from using the proposed modeling approach.

6.3.2 Demand for Wine Futures

We next describe how the demand for wine futures, expressed as $d_f(p_f)$, is developed as a function of futures price p_f. We consider the case when each individual in the futures market has idiosyncratic preferences. In May of calendar year $t + 1$, each individual considers utility from three alternatives: purchasing a wine future, purchasing wine at retail, and no purchase. Individuals discount the value gained in May of calendar year $t + 2$ to the value in May of calendar year $t + 1$ using a risk-adjusted discount rate θ. Because the individuals in the futures market correspond to an affluent population, we consider their decisions to be consistent with risk-neutral decisions; then $\theta = (1+ r)^{-1}$ where r is the risk-free rate over the time-period from May of calendar year $t + 1$ to May of calendar year $t + 2$. The utilities of a random member of the futures market associated with these alternatives are as follows:

$$\begin{aligned}
U_f &= \theta E\left[\tilde{s}_2 | s_1\right] + \varepsilon_f - p_f = \theta s_1 + \varepsilon_f - p_f \\
U_r &= \theta E\left[\tilde{s}_2 | s_1\right] + \varepsilon_r - \theta E\left[p_r\left(\tilde{s}_2\right) | s_1\right] = \varepsilon_r \\
U_0 &= \varepsilon_0
\end{aligned}$$

where ε_f, ε_r, and ε_0 are i.i.d. Gumbel random variables with zero mean and scale parameter β. Gumbel distribution describing the error terms in a multinomial logit model is widely used in literature examining the problems in the retail industry. There are several publications that provide ample empirical support; these include McFadden (2001), Talluri and van Ryzin (2004), and Vulcano et al. (2010).

We first explain the utility expressions U_r and U_0 associated with the retail purchase and no-purchase alternatives. For an individual in the futures market, the value of ε_r is the difference between her valuation of retail wine and the retail price (in the dollar equivalent in May of calendar year $t + 1$). The underlying assumption is that this difference does not depend on the realized bottle score s_2 (or equivalently, due to the deterministic relationship between bottle score and the bottle price). Due to this assumption, each individual knows with certainty her utility (or surplus) from a retail purchase. Similarly, the value of ε_0 is the individual's utility from not purchasing a future or a bottle at retail, which is also known with certainty.

We now turn our attention to the expression for U_f, which is the difference between an individual's valuation of a future and the futures price p_f. An individual's valuation of a future is

$$\theta E\left[\tilde{s}_2 + s_1\right] + \varepsilon_f.$$

The bottle price is uncertain, though its expected value is known as s_1. We describe the market size for wine futures as $M(s_1)$, which is a non-decreasing function of the barrel score s_1 (i.e., $M'(s_1) \geq 0$). A higher barrel score increases the market size as a consequence of the hype it creates. The demand for wine futures can now be determined by the multinomial logit (MNL) model:

$$d_f\left(p_f\right) = M\left(s_1\right) P\left[U_f > \max\left\{U_r, U_0\right\}\right] = M\left(s_1\right)\left[\frac{e^{\left(\theta s_1 - p_f\right)/\beta}}{2 + e^{\left(\theta s_1 - p_f\right)/\beta}}\right]. \quad (6.5)$$

6.4 Analysis

We begin the analysis by establishing the optimal price and quantity coordination in the profit function (6.1) in the absence of the futures quantity constraint (6.4). For a given futures price p_f, the following lemma establishes the optimal futures quantity which is equivalent to the demand established in (6.5); thus, (6.5) also describes the desirable amount of wine that should be allocated as wine futures.

Lemma 1 *For a given futures price p_f, the futures quantity that maximizes (6.1) is* (a) $q_f^* = d_f\left(p_f\right)$ *when* $p_f \geq \phi E\left[p_r\left(\tilde{s}_2 | s_1\right)\right]$; (b) $q_f^* = 0$ *when* $p_f < \phi E\left[p_r\left(\tilde{s}_2 | s_1\right)\right]$.

Lemma 1 indicates that the winemaker sells some of the wine in the form of wine futures as long as the futures price p_f is greater than or equal to the expected retail price discounted to the beginning of the problem, i.e., $q_f^* = d_f\left(p_f\right)$ only when $p_f < \phi E\left[p_r\left(\tilde{s}_2 | s_1\right)\right]$. Otherwise, the winemaker would sell all wine in the retail market, corresponding to Stage 2 of the model.

Aydin and Porteus (2008) show that, in the MNL model, the first-order condition with respect to price yields the optimal price in the absence of a constraint as in (6.4). From equation (6.5), the futures price can be described as a function of demand for wine futures. Alternatively from Lemma 1, the futures price can be expressed in terms of a corresponding optimal futures amount when the winemaker offers a positive amount of futures. Equating $q_f^* = d_f\left(p_f\right)$ and inverting (6.5) provide the following futures price expression for a given futures quantity.

Lemma 2 *For a given futures quantity q_f ($\leq Q$), the futures price that maximizes* (1) *is*

$$p_f^*\left(q_f\right) = \theta s_1 + \beta \ln\left[\frac{M\left(s_1\right) - q_f}{2 q_f}\right] \quad (6.6)$$

Substituting (6.6) into (6.1), the profit function in (6.1) can now be expressed in terms of a single decisions variable.

$$\Pi\left(q_f\right) = \Pi\left(p_f^*\left(q_f\right), q_f\right) = \left\{(\theta - \phi) s_1 + \beta \ln\left[\frac{M\left(s_1\right) - q_f}{2 q_f}\right]\right\} q_f + \phi s_1 Q. \quad (6.7)$$

We substitute (6.7) into the objective function in (6.3), and analyze the constrained MNL model subject to inequality (6.4). We next develop the closed-form expressions for the optimal price, quantity, and profit for the model in (6.3) – (6.4). These expressions utilize the Lambert W function $W(z)$ in Corless et al. (1996) where $W(z)$ describes the value of w satisfying $z = we^w$. Let us define $r^0 = \frac{e^{(\theta-\phi)s_1/\beta - W\left(\frac{e^{(\theta-\phi)s_1/\beta}}{2e}\right)}}{2e + e^{(\theta-\phi)s_1/\beta - W\left(\frac{e^{(\theta-\phi)s_1/\beta}}{2e}\right)}}$ as the optimal proportion of the wine futures market that buys wine in the form of wine futures in the absence of a supply constraint. Recall that the total market size for wine futures consumers is described as $M(s_1)$. In the absence of a supply constraint, then the demand for wine futures is equal to $r^0 \times M(s_1)$.

Proposition 1 *The optimal futures price and futures quantity that maximize* (6.3) *subject to* (6.4) *are*

$$p_f^* = \begin{cases} \phi s_1 + \beta \left[1 + W\left(\frac{e^{(\theta-\phi)s_1/\beta}}{2e}\right)\right] & \text{when } r^0 \leq \frac{Q}{M(s_1)} \\ \theta s_1 + \beta \ln\left[\frac{M(s_1)-Q}{2Q}\right] & \text{when } r^0 > \frac{Q}{M(s_1)} \end{cases} \tag{6.8}$$

$$q_f^* = \begin{cases} M(s_1)\left(\frac{e^{(\theta-\phi)s_1/\beta - W\left(\frac{e^{(\theta-\phi)s_1/\beta}}{2e}\right)}}{2e + e^{(\theta-\phi)s_1/\beta - W\left(\frac{e^{(\theta-\phi)s_1/\beta}}{2e}\right)}}\right) & \text{when } r^0 \leq \frac{Q}{M(s_1)} \\ Q & \text{when } r^0 > \frac{Q}{M(s_1)} \end{cases} \tag{6.9}$$

and the optimal expected profit in (6.3) *is*

$$\Pi^* = \begin{cases} M(s_1)\left[\beta W\left(\frac{e^{(\theta-\phi)s_1/\beta}}{2e}\right) + \phi s_1 \frac{Q}{M(s_1)}\right] & \text{when } r^0 \leq \frac{Q}{M(s_1)} \\ Q\left(\theta s_1 + \beta \ln\left[\frac{M(s_1)-Q}{2Q}\right]\right) & \text{when } r^0 > \frac{Q}{M(s_1)} \end{cases} \tag{6.10}$$

We next provide a sensitivity analysis on the optimal values of the futures price, quantity, and expected profit. We begin the discussion with the influence of the relationship between the consumers' and the winemaker's risk-adjusted discount rates, θ and ϕ, respectively; note that these parameters represent the consumers' and winemaker's risk perception. For example, higher variations in bottle scores imply bigger risk for the winemaker, leading to smaller values of ϕ; thus, the impact of variation in s_2 can be analyzed through decreasing values of ϕ. It is stated earlier that the wine industry is a unique market where the winemakers' risk concern is higher than that of the consumers. Departing from Noparumpa et al. (2015a), we focus on the representative case where $\theta > \phi$.

Higher values of risk aversion for the winemakers is represented with smaller values of ϕ, correspondingly increasing values of $\theta - \phi$. Higher degrees of risk aversion cause the winemaker to allocate a higher percentage of wine for the futures market. This behavior is commonly observed in practice. Small Bordeaux wineries with smaller overall profitability and higher risk concerns (e.g. Evangile, Clos

Fourtet, Troplong Mondot, and Cheval Blanc) allocated more than 25% of their wine as futures on average between 2006 and 2011. During the same time interval, smaller risk winemakers such as Cos d'Estournel and Leoville Poyferre sold less than 15% of their wine on average early in the form of wine futures. The most profitable winemakers with a higher degree of fluctuations in returns, Pavie and Angelus, sold approximately 20% of their wine in the form of futures.

The behavior of the optimal futures price and expected profit, however, are not monotone in ϕ and $\theta - \phi$; and they can feature an increasing or a decreasing behavior depending on the parameter values.

How should a higher barrel score influence the winemaker's allocation of wine to be sold as futures? In this model, the amount of wine allocated for futures (q_f^*) increases in s_1. Similarly, higher barrel scores lead to a higher price of wine futures (p_f^*), and an increased level of optimal expected profit. In this case, the winemaker prefers early cash over the alternative of holding inventory in order to exploit the retail market customers. This result is a consequence of the following two conditions: (1) The assumption of no bias with $E[\tilde{s}_2|s_1] = s_1$, and (2) risk-neutral buyers and risk-averse winemakers, i.e., $\theta > \phi$.

6.5 Empirical Analysis with Bordeaux Winery Data

This section provides an empirical analysis of the stochastic optimization model presented in Sects. 6.3 and 6.4. The analysis considers 12 winemakers from the Bordeaux region, six from the Right Bank and six from the Left Bank wine growing districts. The data used in the analysis is collected from several sources.

Liv-ex is the largest source of fine wine data in the world, and the firm provided all the information regarding the wineries included in the analysis and their futures trades involving futures prices and quantities traded. The wineries included in the analysis are: Angelus (Right Bank), Cheval Blanc (Right Bank), Clos Fourtet (Right Bank), Cos d'Estournel (Left Bank), Ducru Beaucaillou (Left Bank), Duhart Milon (Left Bank), Evangile (Right Bank), Leoville Poyferre (Left Bank), Mission Haut Brion (Left Bank), Pavie (Right Bank), Pichon Lalande (Left Bank) and Troplong Mondot (Right Bank). The data includes the futures of vintages from 2006 to 2011. For the 12 wineries included in the study, there have been 307,909 cases traded in the form of futures in a total of 32,869 futures transactions.

Barrel and bottle scores indicate the quality of the wine. The data for the barrel and bottle scores, and the production quantities, are collected from the most influential wine magazine, *Wine Spectator*.

6.5.1 *Wine Futures as a Quantity Lever*

This stochastic optimization model uses wine futures as a quantity lever as the winemaker sells a proportion of the wine in the form of wine futures. In this section, we develop a regression model in order to predict the percentage of wine allocated as futures based on *Wine Spectator*'s barrel scores. The regression analysis helps demonstrate the robustness of this optimization model.

We describe the percentage of wine allocated as futures with r_{jt} and barrel scores with s_{1jt} for winery j and vintage t, the mean and standard deviations of the percentage of wine sold as futures and barrel scores with $\bar{r}_j$, σ_{α_j} and the mean and standard deviations of barrel scores with $\bar{s}_{1j}$ and $\sigma_{s_{1j}}$, respectively. We normalize the values of the ratio of wine allocated as futures and barrel scores and describe them as follows: $\widehat{r}_{jt} = \frac{r_{jt}-\bar{r}_j}{\sigma_{r_j}}$ and $\widehat{s}_{1jt} = \frac{s_{1jt}-\bar{s}_{1j}}{\sigma_{s_{1j}}}$.

The first analysis regresses the normalized values of percentage of wine allocated as futures $(\widehat{r}_{jt})$ based on the normalized values of Wine Spectator's barrel tasting scores $(\widehat{s}_{1jt})$. Table 6.1 provides the results of this regression analysis, and shows that barrel score is a statistically significant variable at less than 1%. Thus, we conclude that barrel score explains a fairly large portion of the amount of wine that should be allocated as wine futures. Figure 6.3 shows how well this regression model fits between the actual and forecasted percentage allocation.

Figure 6.4 demonstrates the impact of barrel scores on each winery's percentage allocation decision during each vintage. For each vintage, given the Wine Spectator barrel score, it shows each winery's actual allocation (labeled as "Futures Allocation") and the percentage of wine that should have been allocated according to the statistical analysis reported in Table 6.1 (labeled as "Forecast Futures Allocation").

Table 6.1 Summary of regression results for the normalized values of wine allocated as futures versus the normalized values of barrel scores

Parameter	Coefficient	(*p-value*)
Intercept	3.48×10^{-16}	(1)
Barrel score $(\widehat{s}_{1jt})$	0.73	(2.16×10^{-13})[a]
Adjusted R^2	0.53	

[a]Implies that the variable is significant at 0.01 level

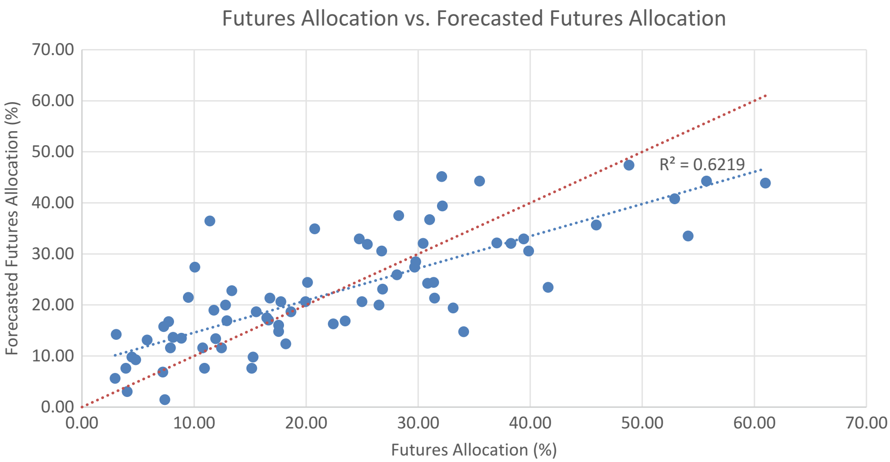


Fig. 6.3 The fit between the normalized actual and forecasted futures allocation

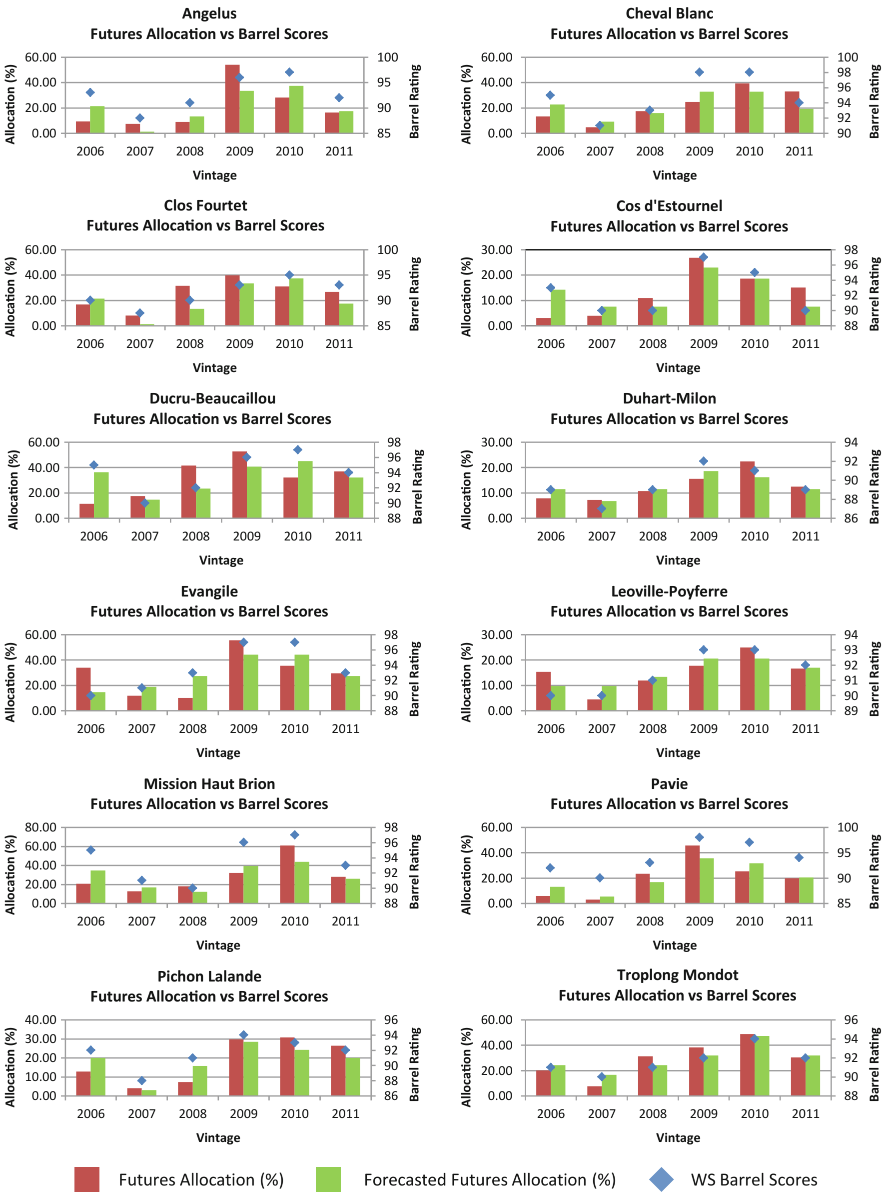


Fig. 6.4 Actual and forecasted percentage of wine allocated for futures predicted from the regression model vs. barrel scores from *Wine Spectator*

Figure 6.5 shows, during each vintage, how each winemaker allocated its wine for the futures market and what the regression-based model suggested as the percentage to be sold as futures based on the barrel scores.

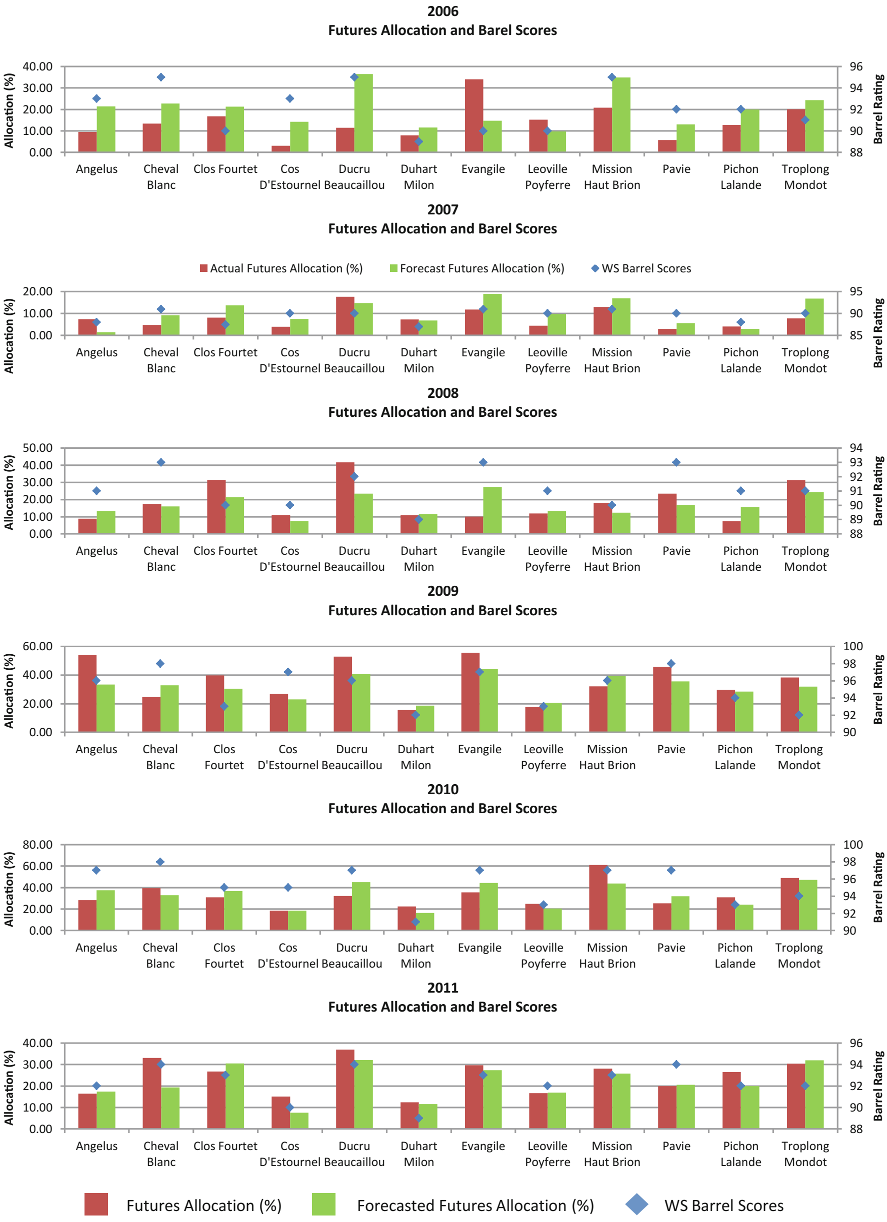


Fig. 6.5 Actual and forecasted futures allocation percentage of Bordeaux wineries vs. barrel scores during each vintage

6.5.2 *Wine Futures as a Price Lever*

This stochastic optimization model describes that wine futures can be used as a price lever in order to mitigate quality uncertainty stemming from the randomness in bottle reviews of the tasting experts. In this section, we develop a regression model in order to predict the futures price using *Wine Spectator*'s barrel scores. The regression analysis once again helps demonstrate the robustness of this optimization model.

Using the same 12 wineries, we denote the futures price for winery j and its vintage t from 2006 to 2011 by f_{jt}, the mean and standard deviations of the futures prices for winery j with $\bar{f}_j$ and σ_{f_j}, respectively, and the normalized futures price with $\widehat{f_{jt}} = \frac{f_{jt} - \bar{f}_j}{\sigma_{f_j}}$.

The results of the regression analysis is presented in Table 6.2. It shows that the barrel score is a statistically significant variable at less than 1%. The adjusted R^2 of 0.70 indicates that the barrel score explains a fairly large portion of the decision regarding futures prices. Figure 6.6 shows the fit of this regression model by comparing the actual futures price with the forecasted futures price.

Figure 6.7 demonstrates the impact of barrel scores on each winery's futures price decision during each vintage. For each vintage, given the *Wine Spectator* barrel

Table 6.2 Summary of regression results for the normalized values of futures prices versus the normalized values of barrel scores

Parameter	Coefficient	(*p-value*)
Intercept	3.63×10^{-16}	(1)
Barrel score ($\widehat{s}_{1jt}$)	0.841	$(2.18 \times 10^{-20})^{a}$
Adjusted R^2	0.70	

[a]Implies that the variable is significant at 0.01 level

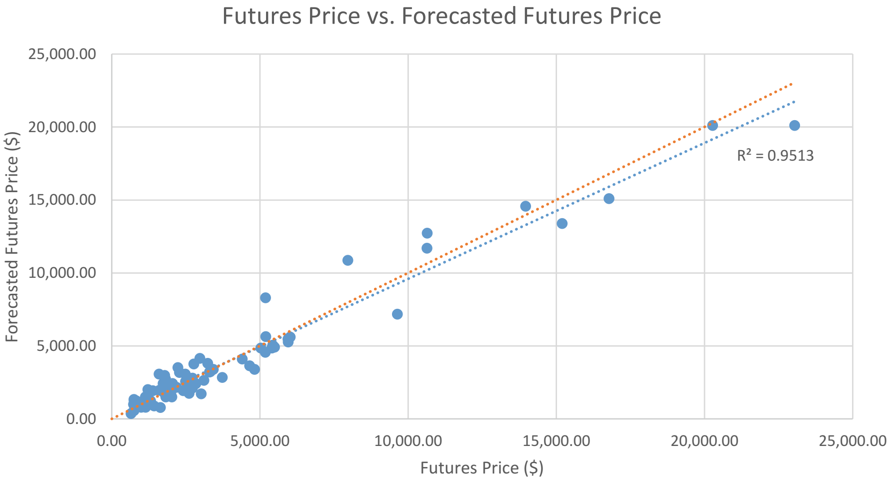


Fig. 6.6 The fit between the normalized actual and forecasted futures prices

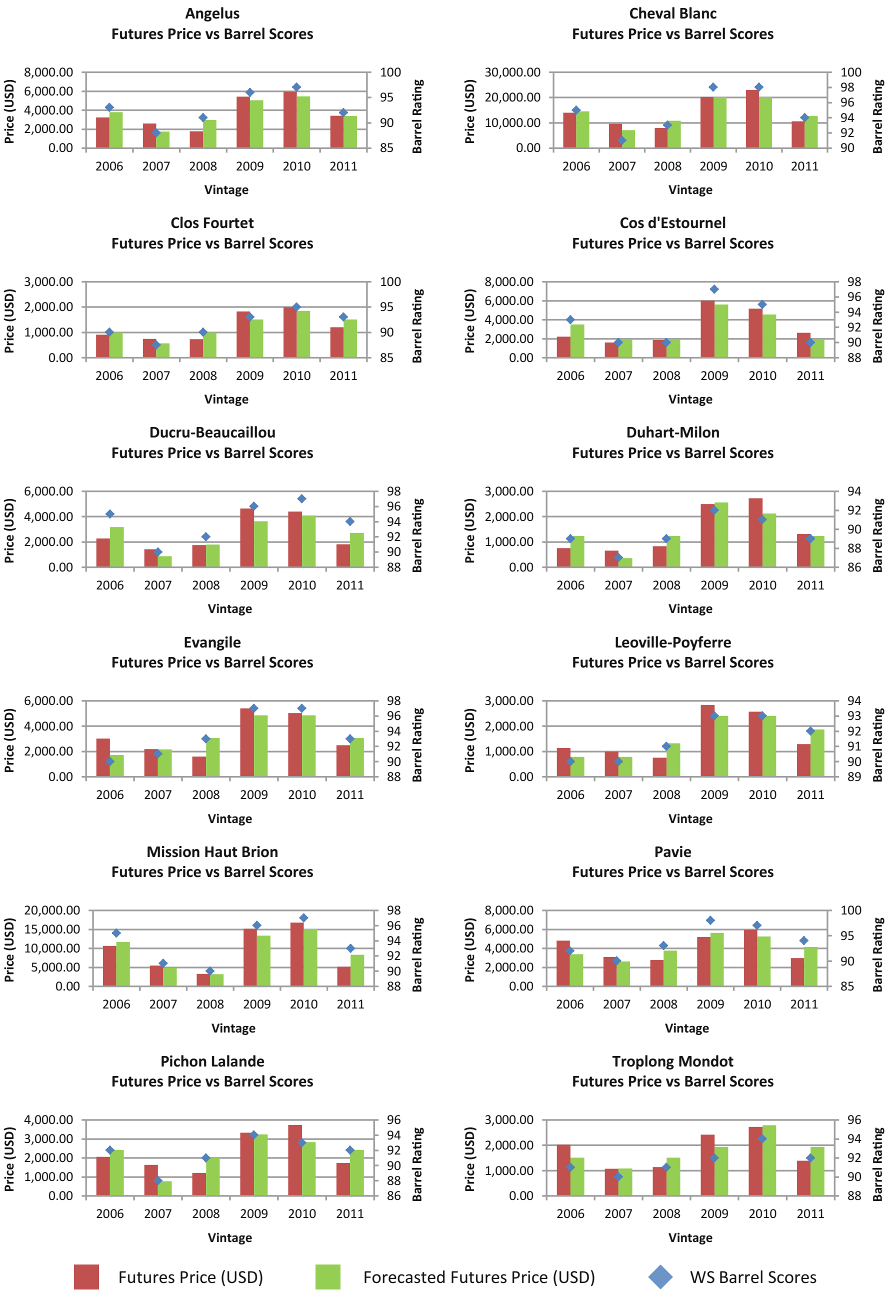


Fig. 6.7 Actual futures price of wine, and forecasted futures price predicted from the regression model vs. barrel scores from *Wine Spectator*

Table 6.3 The correlation coefficients barrel scores, futures allocation percentage, forecasted futures allocation percentage, futures price, and forecasted futures price

	Barrel score (s_1)	Allocation (r)	Futures price (p_f)	Forecast allocation (r_{jt})	Forecast futures price (f_{jt})
Barrel score (s_1)	1				
Allocation (r)	0.644	1			
Futures price (p_f)	0.621	0.326	1		
Forecast allocation (r_{jt})	0.824	0.789	0.380	1	
Forecast futures price (f_{jt})	0.634	0.303	0.975	0.384	1

score, it shows each winery's actual futures price decision (labeled as "Futures Price") and the futures price forecasted based on the statistical analysis reported in Table 6.2 (labeled as "Forecast Futures Price").

Table 6.3 presents the correlation coefficient values between the Wine Spectator barrel scores (s_1), the allocation percentages assigned by winemakers (r), the futures price (p_f), and the forecasts using the regression models for the allocation (r_{jt}) and for the futures price (f_{jt}) decisions. Table 6.3 demonstrates that the barrel score (s_1) shows a 64% positive correlation with the winemaker's allocation decision (r) and a 62.1% positive correlation with the futures decision (p_f). These correlation values are significant. While the percentage of wine allocated for the futures market and the futures price exhibit relatively lower correlation (32.6%), this statistical analysis leads to strong fit with a 78.9% positive correlation between the amount of wine allocated for the futures market and its estimate. We obtain a strong fit for the futures price using *Wine Spectator*'s barrel scores, and this can be viewed from 97.5% positive correlation between the futures price and its estimate. As a consequence, we conclude that there is a strong relationship between the barrel scores, the percentage of wine allocated for the futures market and the futures price. In conclusion, the statistical analysis provides for the foundation for using barrel scores in the stochastic optimization model.

6.5.3 Financial Benefit from the Proposed Stochastic Optimization Model

The analysis in this section compares the firm's actual allocation quantity for wine futures and the futures price (p_f^0, q_f^0) provided by Liv-ex.com with the optimal choices (p_f^*, q_f^*) developed from solving the model presented in (6.3) – (6.4).

Table 6.4 presents the financial benefit from utilizing the stochastic optimization model in the 12 Bordeaux winemakers during the 2006 and 2011 vintages. In the empirical analysis, we estimate consumers' valuation of fine wine using the risk-free rate in order to reflect the structure of this market that is populated with the affluent customer base. We describe consumer's valuation with $\theta = (1 + r_f)^{-1}$ where r_f is

Table 6.4 The financial benefit from the stochastic optimization model with $\theta = 0.97561$; $\beta = 24$

Winemaker	ϕ	Minimum $\Delta\Pi$	Maximum $\Delta\Pi$	Average $\Delta\Pi$
Angelus	0.96936308	1.83	9.64	4.75
Cheval Blanc	0.86918809	1.34	10.59	5.18
Clos Fourtet	0.88701179	1.91	6.98	5.29
Cos d'Estournel	0.87673835	0.88	4.72	3.01
Ducru Beaucaillo	0.88961788	2.44	9.11	5.94
Duhart Milon	0.79816123	1.06	2.51	1.58
Evangile	0.85688923	2.74	11.69	7.57
Leoville Poyferre	0.9082983	0.47	2.29	1.47
Mission Haut Brion	0.94221522	2.76	8.22	4.55
Pavie	0.97247639	0.62	8.88	4.04
Pichon Lalande	0.84258235	0.74	6.71	3.26
Troplong Mondot	0.83791897	0.69	4.08	2.54
Weighted average		**1.46**	**7.12**	**4.10**

the risk-free rate. Using the European Central Bank interest rate of 0.025, we have $\theta = 0.97561$.

We estimate the winemaker's preference, the risk-adjusted discount rate ϕ, using a Capital Asset Pricing Model (CAPM) perspective. We describe the winemaker's preference as $\phi = (1+ r_f + \gamma\ (r_m - r_f))^{-1}$ where $r_{m.}$ is the market return, $r_{m.} - r_f$ is the risk premium, and γ is the winemaker's risk measure following the Capital Asset Pricing Model (CAPM) approach. We estimate $r_{m.} = 0.104262504$ through the average annual market returns in the Liv-ex 100 index from 2006 to 2013 describing the most sought after 100 wines in the world; it is important to note that the vintages of the most sought after 100 wines do not exist in our data set. We estimate each winemaker's risk premium through the covariance between the returns of the specific winemaker and the market returns (defined as $COV(r_j, r_m)$) divided by the variance in market returns (defined as $VAR(r_m)$), i.e., $\gamma = COV(r_j, r_m)/VAR(r_m)$. Market size is also provided by Liv-ex. We estimate the consumer heterogeneity parameter β in the Gumbel distribution by inverting the Bordeaux winemaker's allocation percentages; because most of these wineries cater to the same variety of consumer, we use the average value corresponding to $\beta = 24$.

Table 6.4 demonstrates that the proposed stochastic optimization model is expected to improve profits of the 12 Bordeaux winemakers by 4.10%. On average, the minimum improvement is 1.46% and the maximum is 7.12%; and the overall minimum is 0.62% and the overall maximum is 11.69%.

In sum, we conclude that the proposed stochastic optimization model is financially beneficial for the Bordeaux winemakers in determining their optimal futures price and futures allocation as it can improve their profits by more than 4%.

6.5.4 Financial Impact from a Wine Futures Market

We next present the benefit from establishing a futures market for winemakers. As stated earlier, the US wine industry does not have a futures market. What would be the benefit from establishing a wine futures market in the US? While this is hard to demonstrate in precision, we attempt in identifying potential benefits by comparing the optimal profit obtained from solving the proposed stochastic optimization in in (6.3) – (6.4) with the optimal profit that can obtained from setting the futures quantity equal to zero, i.e., $q_f = 0$. Note that when there is no futures market, the firm has to sell all of its wine in the retail market; we describe the profit that can be obtained in the absence of a futures market by Π^0. The profit in the absence of a futures market is calculated by substituting $q_f = 0$ in (6.1); this provides $\Pi^0 = \phi s_1 Q$. The percentage impact of wine futures on the profit of the winemaker is then described as follows:

$$\Delta\Pi^0 = \left(\Pi\left(p_f^*, q_f^*\right) - \Pi^0\right) / \Pi^0 \times 100\%. \tag{6.11}$$

The directional impact of a higher barrel score (s_1) on $\Delta\Pi^0$ is not monotone, and is parameter-dependent. We can observe from (6.11) that $\Delta\Pi^0$ is higher for a highly risk-averse winemaker with smaller values of ϕ.

Table 6.5 demonstrates the financial benefit that can be obtained from having a futures market. Using the 2006 – 2011 vintages of the 12 wineries employed in the empirical analysis, and it estimates the financial benefit from having a futures market to be 7.82% on average. The average percentage improvement in profits from the presence of a futures market ranges from 3.19% to 17.19%. The minimum financial benefit occurs at the low barrel scores as observed at Leoville Poyferre with a 1.08% profit improvement; the highest benefit is observed with high barrel scores at Cheval Blanc with a 23.57% profit improvement. It is important to note here that Table 6.5 shows that the financial benefit from the presence of a futures market for the Bordeaux wineries is 7.82%, and this number is smaller than the estimate of 10.10% provided in Noparumpa et al. (2015a, b). There are two differences in our study when compared with Noparumpa et al. (2015a, b). First, this study considers risk-neutral consumers, rather than risk-averse consumers, reflecting the operating environment in the wine industry. Thus, we believe that this is a better estimate representing the wine futures market. Second, we use barrel and bottle scores established by *Wine Spectator*, the most widely distributed magazine in the wine industry, and Noparumpa et al. (2015a, b) relies on Robert Parker scores. Table 6.6 provides the correlation between the barrel scores and bottle scores establishes by Robert Parker and *Wine Spectator* for the 12 wineries used in the earlier analysis during the same vintages of 2006 – 2011. It is evident that there is strong correlation between the tasting expert reviews, the correlation coefficient between the barrel scores of Robert Parker and *Wine Spectator* for the 12 winemakers in this study is 80.3% during the 2006 – 2011 vintages. In conclusion, Table 6.5 demonstrates that the wine futures market creates a significant financial benefit to the Bordeaux winemakers.

Table 6.5 The financial benefit from the presence of a wine futures market in winemaker profits with $\theta = 0.97561$; $\beta = 24$

Winemaker	Minimum r	Maximum r	Average r	Minimum $\Delta\Pi^0$	Maximum $\Delta\Pi^0$	Average $\Delta\Pi^0$
Angelus	9.88	62.29	27.83	3.23	18.65	8.61
Cheval Blanc	7.58	68.20	32.57	2.81	23.57	11.58
Clos Fourtet	11.05	48.88	34.07	4.12	17.21	12.19
Cos d'Estournel	4.98	32.76	18.68	1.79	11.28	6.73
Ducru Beaucaillo	15.58	66.90	40.54	5.35	22.75	14.09
Duhart Milon	8.23	23.13	13.45	3.61	9.74	5.75
Evangile	16.10	84.20	56.50	5.97	30.02	17.19
Leoville Poyferre	3.07	14.79	9.23	1.08	5.03	3.19
Mission Haut Brion	16.29	58.23	29.52	5.36	17.99	9.42
Pavie	3.46	57.11	27.03	1.10	16.67	7.74
Pichon Lalande	4.83	39.63	22.25	1.93	14.93	8.59
Troplong Mondot	6.17	34.21	22.34	2.44	12.99	9.06
Weighted average			**21.90**			**7.82**

Table 6.6 The correlation coefficients between the barrel scores and the bottle scores established by Robert Parker and *Wine Spectator* for the same 12 wineries used in this study during the 2006 – 2011 vintages

	Robert Parker barrel score	Robert Parker bottle score	*Wine Spectator* barrel score	*Wine Spectator* bottle score
Robert Parker barrel score	1			
Robert Parker bottle score	0.905	1		
Wine Spectator barrel score	0.803	0.774	1	
Wine Spectator bottle score	0.752	0.731	0.860	1

Table 6.5 also demonstrates how futures quantity is a beneficial lever in mitigating the winemaker's quality uncertainty. The analysis shows that these wineries would benefit by using wine futures as a quantity lever: They should allocate on average 21.90% of their wine as futures, with a minimum of 9.23% and a maximum of 56.50% on average. If they get low barrel scores, the 12 winemakers allocate less wine for wine futures; the minimum occurs at Leoville Poyferre with a 3.07% of wine dedicated to wine futures. High barrel scores are desirable, and when a winery receives a high barrel score, it can reserves more wine for the futures market. This is exemplified in Evangile who allocated 84.20% of its production up for sale in the form of wine futures. We conclude that selling wine while aging in the barrel in the form of wine futures provides a good quantity lever to these winemakers interested in reducing the negative consequences of bottle score uncertainty.

Table 6.7 The financial benefit from the presence of a wine futures market at Heart & Hands Wine Co. using parameters $\theta = 0.998801$; $\phi = 0.76595$, $\beta = 10$

Varietal	Vintage	r	$\Delta\Pi^0$
Pinot noir barrel reserve	2007	45.96	11.91
	2008	37.41	12.55
	2009	45.03	11.33
	2010	46.89	11.93
Riesling	2008	55.92	14.15
	2009	59.11	14.79
	2010	55.27	13.90
	2011	59.60	14.99
Average		**47.61**	**12.51**

While Liv-ex is a beneficial electronic exchange platform for trading wine, many small and artisanal winemakers cannot benefit from the presence of this market. Wineries that are traded in Liv-ex are established winemakers with a recognized brand name and image. We argue that artisanal and boutique wineries would particularly benefit from establishing a futures market. In the US, most winemakers are small and do not possess the brand recognition of Bordeaux winemakers. Similarly, few Italian winemakers are traded in Liv-ex, and a majority of winemakers in this country have limited resources to negate the financial consequences from poor reviews. We expect these small and artisanal winemakers to allocate a higher percentage of their wine to be sold in the form of wine futures; specifically, the quantity lever would be used significantly. Similarly, we expect them to reduce their futures price significantly, and therefore, they would engage in using futures as a price lever. These arguments are demonstrated in Table 6.7 through the analysis of Heart & Hands Wine Co., a small and artisanal winemaker in the Finger Lakes region in the State of New York. Heart & Hands Wine Co. is gaining significant reputation for its stellar Pinot Noir and is in the process of becoming a popular winemaker. We estimate the US consumers' valuation by using the risk-free rate of return based on the 12-month US Treasury Bond; we have $r_f = 0.0012$, leading to $\theta = (1 + r_f)^{-1} = 0.998801$. We again utilize the CAPM approach in order to estimate the winemaker's risk preference; we have $\phi = (1 + r_f + \gamma\,(r_m - r_f))^{-1} = 0.76595$. Because consumers in the US, and particularly for this small winemaker, are more homogenous compared to the Bordeaux winemakers, we describe consumer heterogeneity through a Gumbel distribution with mean of zero and a smaller dispersion parameter at $\beta = 10$. We use the scores established by *Wine Spectator* in this analysis; it is also important to note that Robert Parker and The Wine Advocate does not provide reviews of the small and artisanal winemakers such Heart & Hands Wine Co.

Table 6.7 demonstrates that wine futures offer financially more beneficial price and quantity levers for the small and artisanal winemakers than the Bordeaux wineries. Heart & Hands Wine Co. improves its profit by 12.51% on average with a minimum financial benefit of 11.33% and a maximum financial benefit of 14.99%. This winery has consistently lower scores than Bordeaux wineries, and the proposed model recommends to allocate a significantly larger percentage of its wine as futures: 47.61%. Thus, the quantity lever of wine futures is an extremely important risk mitigation tool for small and boutique winemakers.

6.6 Conclusions

This paper develops a stochastic optimization model and establishes the use of wine futures as a price and quantity lever in order to mitigate quality uncertainty. The winemaker financially benefits from the use of wine futures, but more importantly, reduces the negative consequences of uncertain tasting expert reviews that get established when the wine is bottled. Selling some of her wine in advance, the winemaker recuperates her cash investment in a liquid that is uncertain in value; the firm can use this money to reinvest in business, improve quality, and expand her growth initiatives.

The study makes two sets of contributions. First, we develop an analytical model that helps winemakers improve their profits. The proposed model incorporates uncertain consumer valuations of wine futures and bottled wine and the random bottle score that is assigned to the wine at the end of the production process. The analysis leads to closed-form expressions for the optimal futures price, futures quantity and the expected profit.

Second, we test the model by illustrating how it benefits the winemakers. We show that the proposed stochastic model can improve the profits of Bordeaux winemakers by 4.10% on average. We also estimate the financial benefits from using the futures market for these Bordeaux winemakers: The futures market helps improve their profits by 7.82% on average. Thus, the model makes a substantial contribution to their bottom line profits. Finally, establishing a futures market in other regions, e.g. the US and Italy, can be extremely beneficial for the small and artisanal winemakers. Using one small winery from the Finger Lakes region in the US, we demonstrate that this small winemaker would sell a higher percentage of their wine with deeper discounts benefiting her more than the Bordeaux winemakers. Thus, establishing a futures market would enable small and artisanal winemakers to utilize these price and quantity levers to create a healthy and sustainable growth opportunity.

This study also sheds light into the benefits from price efficiency over the traditional practice of market-clearing price mechanisms. It is often believed that winemakers, as well as many retailers, use market-clearing prices in order to sell out the inventory of short selling season items. Wine for a specific vintage can be perceived as a short selling season item, because winemakers need to replace the shelf space and limited storage space (for barrels dedicated to aging the wine) with the upcoming vintage's bottles and barrels. This paper demonstrates that, by using price as a lever, winemakers can increase their expected profits.

Acknowledgments We are grateful to Liv-ex.com for their generosity with data; the study would not have been possible without their contribution. We are grateful to Mr. Ben O'Donnell of *Wine Spectator* for his insights on our study. We are also thankful to Tom and Susan Higgins of Heart and Hands Wine Co. for sharing data about the winery and the industry. The paper has benefited from participants' feedback through presentations at Arizona State University, the Third Supply Chain Finance Conference in 2013 in Eindhoven, The Netherlands, at Analytics Operations Engineering Inc., at Boston University, at the 2014 Euro Working Group for Commodities and Financial Modelling in Milan, Italy, and at the University of Milano Bicocca. This study was partially supported by the Robert H. Brethen Operations Management Institute and the H.H. Franklin Center for Supply Chain Management at Syracuse University.

Appendix

Proof of Lemma 1 Taking the derivative of (6.1) with respect to q_f provides the result.

$$\partial\Pi\left(p_f, q_f\right)/\partial q_f = \begin{cases} p_f - \phi E\left[p_r\left(\tilde{s}_2\left|s_1\right.\right)\right] & \text{if } q_f < d_f\left(p_f\right) \\ 0 & \text{if } q_f \geq d_f\left(p_f\right) \end{cases}$$

Because $E\left[\tilde{s}_2|s_1\right] = s_1$, the expected retail price in May of calendar year $t+1$ is equal to the barrel score, i.e., $E\left[p_r\left(\tilde{s}_2|s_1\right)\right] = s_1$. When $p_f \geq \phi E\left[p_r\left(\tilde{s}_2|s_1\right)\right]$, the derivative is positive for $q_f < d_f(p_f)$, and is equal to zero for $q_f \geq d_f(p_f)$. Thus, increasing q_f to $d_f(p_f)$ provides a positive improvement in the expected profit. When $p_f \geq \phi E\left[p_r\left(\tilde{s}_2|s_1\right)\right]$, the winemaker sells all of the wine in Stage 2 in the retail market. □

Proof of Lemma 2 We have $E\left[p_r\left(\tilde{s}_2|s_1\right)\right] = s_1$, and we take the natural log of $q_f = d_f(p_f)$ where $d_f(p_f)$ is expressed as in (6.5). Thus,

$$\begin{aligned} q_f = d_f\left(p_f\right) &= M\left(s_1\right)\left[\frac{e^{\left(\theta s_1 - p_f\right)/\beta}}{2 + e^{\left(\theta s_1 - p_f\right)/\beta}}\right] \Rightarrow \frac{q_f}{M\left(s_1\right)}\left(2 + e^{\left(\theta s_1 - p_f\right)/\beta}\right) \\ &= e^{\left(\theta s_1 - p_f\right)/\beta} \Rightarrow \frac{2q_f}{M\left(s_1\right) - 2q_f} = e^{\left(\theta s_1 - p_f\right)/\beta}. \end{aligned}$$

Taking the natural logarithm of both sides provides:

$$\ln\left[\frac{2q_f}{M\left(s_1\right) - 2q_f}\right] = \beta\left(\theta s_1 - p_f\right)$$

Rearranging the terms, we obtain the futures price expression in (6.6).

$$p_f\left(q_f\right) = \theta s_1 - \beta\ln\left[\frac{2q_f}{M\left(s_1\right) - q_f}\right] = \theta s_1 + \beta\ln\left[\frac{M\left(s_1\right) - q_f}{2q}\right].$$

□

Lemma A1 *Maximizing the objective function in* (6.3) *is equivalent to maximizing the expected profit expression in* (6.1).

Proof of Lemma A1 The objective function in (6.3) can be rewritten as follows:

$$\Delta\Pi\left(p_f, q_f\right) = \left(1/\,\Pi\left(p_f^0, q_f^0\right)\right)\left(\Pi\left(p_f, q_f\right) - \Pi\left(p_f^0, q_f^0\right)\right) \times 100\%$$

Because the values of p_f^0 and q_f^0 are given, the expected profit expression for the winemaker's current profit level, described with $\Pi\left(p_f^0, q_f^0\right)$, is constant. Thus, maximizing $\Delta\Pi(p_f, q_f)$ is equivalent to maximizing the expected profit expression $\Pi(p_f, q_f)$ in (6.1). □

Proof of Proposition 1 From Lemma A1, we know that maximizing $\Delta\Pi(p_f, q_f)$ in (6.3) is equivalent to maximizing the expected profit $\Pi(p_f, q_f)$ in (6.1). Thus, we focus on the properties of (6.1). Moreover, we know that $\Pi(p_f, q_f)$ can be expressed as a single decision variable function as in (6.7). Thus, it is sufficient to show that $\Pi(q_f)$ is concave in q_f. Following the proof of Theorem 1 in Li and Huh (2011); it can be shown that $\Pi''(q_f) < 0$. Using the first-order condition and (6.6), we have the futures price can be expressed as follows:

$$p_f\left(q_f\right) = \phi s_1 + \frac{\beta M\left(s_1\right)}{M\left(s_1\right) - q_f} = \beta + \phi s_1 + \frac{\beta q_f}{M\left(s_1\right) - q_f}.$$

Using the approach described in the derivations of Proposition 1 of Noparumpa et al. (2015a), the optimal unconstrained futures quantity can be obtained as follows:

$$q_f^0 = M\left(s_1\right)\left(\frac{e^{(\theta-\phi)s_1/\beta - W\left(\frac{e^{(\theta-\phi)s_1/\beta}}{2e}\right)}}{2e + e^{(\theta-\phi)s_1/\beta - W\left(\frac{e^{(\theta-\phi)s_1/\beta}}{2e}\right)}}\right)$$

Iff $q_f^0 \leq Q$, then $q_f^* = q_f^0$, and the optimal profit is equivalent to the unconstrained optimal profit, and

$$p_f^* = \phi s_1 + \beta\left[1 + W\left(\frac{e^{(\theta-\phi)s_1/\beta}}{2e}\right)\right]$$

If $q_f^0 \geq Q$, then the supply constraint is binding, i.e., $q_f^* = Q$. In this case, the optimal price is obtained by substituting $q_f^* = Q$ in (6.6), and the optimal profit is obtained by substituting the revised price expression into (6.1). □

References

H.H. Ali, S. Lecocq, M. Visser, The impact of gurus: Parker grades and en primeur wine prices. Econ. J. **118**(529), F158–F173 (2008)

O. Ashenfelter, Prediciting the quality and prices of Bordeaux wine. J. Wine Econ. **5**(1), 40–52 (2010)

O. Ashenfelter, G.V. Jones, The demand for expert opinion: Bordeaux wine. J. Wine Econ. **8**(3), 285–293 (2013)

O. Ashenfelter, D. Ashmore, R. Lalonde, Bordeaux wine vintage quality and the weather. Chance **8**(4), 7–14 (1995)

G. Aydin, E.L. Porteus, Joint inventory and pricing decisions for an assortment. Oper. Res. **56**(5), 1247–1255 (2008)

J.C. Bodington, Evaluating wine-tasting results and randomness with a mixture of rank preference models. J. Wine Econ. **10**(1), 31–46 (2015)

T. Boyacı, Ö. Özer, Information acquisition for capacity planning via pricing and advance selling: When to stop and act? Oper. Res. **58**(5), 1328–1349 (2010)

S.H. Cho, C.S. Tang, Advance selling in a supply chain under uncertain supply and demand. Manuf. Serv. Oper. Manag. **15**(2), 305–319 (2013)

R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey, D.E. Knuth, On the Lambert W function. Adv. Comput. Math. **5**(1), 329–359 (1996)

J.D. Dana, N.C. Petruzzi, Note: The newsvendor model with endogenous demand. Manag. Sci. **47**(11), 1488–1497 (2001)

S. Fay, J.H. Xie, The economics of buyer uncertainty: Advance selling vs. probabilistic selling. Mark. Sci. **29**(6), 1040–1056 (2010)

A. Federgruen, A. Heching, Combined pricing and inventory control under uncertainty. Oper. Res. **47**(3), 454–475 (1999)

A. Federgruen, A. Heching, Multilocation combined pricing and inventory control. Manuf. Serv. Oper. Manag. **4**(4), 275–295 (2002)

I.L. Gale, T.J. Holmes, The efficiency of advance-purchase discounts in the presence of aggregate demand uncertainty. Int. J. Ind. Organ. **10**(3), 413–437 (1992)

I.L. Gale, T.J. Holmes, Advance-purchase discounts and monopoly allocation of capacity. Am. Econ. Rev. **83**(1), 135–146 (1993)

S. Gheibi, B. Kazaz, S. Webster, *Direct-trade and the third-wave coffee: Sourcing and pricing a specialty product under uncertainty* (Syracuse University, Working Paper, 2017)

M.H. Hekimoğlu, B. Kazaz, S. Webster, Wine analytics: Fine wine pricing and selection under weather and market uncertainty. Manufact. Ser. Operat. Manag. **19**(2), 202–215 (2017)

P.C. Jones, T. Lowe, R.D. Traub, G. Keller, Matching supply and demand: The value of a second chance in producing hybrid seed corn. Manuf. Serv. Oper. Manag. **3**(2), 122–137 (2001)

B. Kazaz, Production planning under yield and demand uncertainty. Manuf. Serv. Oper. Manag. **6**(3), 209–224 (2004)

B. Kazaz, S. Webster, The impact of yield dependent trading costs on pricing and production planning under supply uncertainty. Manuf. Serv. Oper. Manag. **13**(3), 404–417 (2011)

B. Kazaz, S. Webster. Technical note – price-setting newsvendor problems with uncertain supply and risk aversion. Forthcoming at *Operations Research* (2015)

A. Kocabıyıkoğlu, I. Popescu, An elasticity approach to the newsvendor with price-sensitive demand. Oper. Res. **59**(2), 301–312 (2011)

H. Li, W.T. Huh, Pricing multiple products with the multinomial logit and nested logit concavity: Concavity and implications. Manuf. Serv. Oper. Manag. **13**(4), 549–563 (2011)

P. Masset, J.P. Weisskopf, M. Cossutta, Wine tasters, ratings and en primeur prices. J. Wine Econ. **10**(1), 75–107 (2015)

D. McFadden, Economic choices. Am. Econ. Rev. **91**(3), 351–378 (2001)

T. Noparumpa, B. Kazaz, S. Webster. Wine futures and advance selling under quality uncertainty. Manuf. Serv. Oper. Manag. **17**(3), 411–426 (2015a)

T. Noparumpa, B. Kazaz, S. Webster. Production planning under supply and quality uncertainty with two customer segments and downward substitution. Working Paper (2015b)

I. Olkin, Y. Lou, L. Stokes, J. Cao, Analyses of wine-tasting data: A tutorial. J. Wine Econ. **10**(1), 4–30 (2015)

N. Petruzzi, M. Dada, Pricing and the newsvendor problem: A review with extension. Oper. Res. **47**(2), 183–194 (1999)

S.M. Shugan, J.H. Xie, Advance pricing of services and other implications of separating purchase and consumption. J. Serv. Res. **2**(3), 227–239 (2000)

S.M. Shugan, J.H. Xie, Advance selling as a competitive marketing tool. J. Res. Mark. **22**(3), 351–373 (2005)

E.T. Stuen, J.R. Miller, R.W. Stone, An analysis of wine critic consensus: A study of Washington and California wines. J. Wine Econ. **10**(1), 47–61 (2015)

K. Talluri, G. Van Ryzin, *The Theory and Practice of Revenue Management* (Kluwer Academic, Boston, 2004)

C.S. Tang, W.S. Lim, Advance selling in the presence of speculators and forward looking consumers. Prod. Oper. Manag. **22**(3), 571–587 (2013)

J.A. Van Mieghem, M. Dada, Price versus production postponement: Capacity and competition. Manag. Sci. **45**(12), 1631–1649 (1999)

G. Vulcano, G. Van Ryzin, W. Chaar, Choice-based revenue management: An empirical study of estimation and optimization. Manufact. Ser. Operat. Manag. **12**(3), 371–392 (2010)

J. Xie, S.M. Shugan, Electronic tickets, smart cards, and online payments: When and how to advance sell. Mark. Sci. **20**(3), 219–243 (2001)

Chapter 7
VIX Computation Based on Affine Stochastic Volatility Models in Discrete Time

A. Hitaj, L. Mercuri, and E. Rroji

Abstract We propose a class of discrete-time stochastic volatility models that, in a parsimonious way, capture the time-varying higher moments observed in financial series. Three desirable results are obtained. First, we have a recursive procedure for the log-price characteristic function which allows a semi-analytical formula for option prices as in Heston and Nandi (Rev Financ Stud 13(3):585–625, 2000). Second, we reproduce some features of the VIX Index. Finally, we derive a simple formula for the VIX index and use it for option pricing.

Keywords Affine stochastic volatility • VIX • Implied volatility surface

7.1 Introduction

The Black and Scholes model (see Black and Scholes, 1973) is probably the most famous model proposed for option pricing. Despite its success, the drawbacks in representing real market stylized facts are well documented by an increasing empirical literature (see Embrechts et al., 1997, and the references therein). Since Mandelbrot (1963), empirical results have shown that the process describing log returns is not a Brownian motion. Indeed, financial time series exhibit heavy tails, asymmetric distribution, persistence and clustering in volatility.

A. Hitaj (✉)
Department of Quantitative Methods, University of Milano-Bicocca, Milan, Italy
e-mail: asmerilda.hitaj1@unimib.it

L. Mercuri
Department of Economics, Management and Quantitative Methods,
University of Milan, Milan, Italy

CREST Japan Science and Technology Agency, Kawaguchi, Japan
e-mail: lorenzo.mercuri@unimi.it

E. Rroji
Department of Economics, Business, Mathematical and Statistical Sciences,
University of Trieste, Trieste, Italy
e-mail: erroji@units.it

G. Consigli et al. (eds.), *Handbook of Recent Advances in Commodity and Financial Modeling*, International Series in Operations Research & Management Science 257,
DOI 10.1007/978-3-319-61320-8_7

Several models have been proposed in continuous and discrete time. A first improvement is obtained through the introduction of the Lévy processes with jumps in finance. For instance, Merton (1976) introduced a Jump diffusion model in the evaluation of option prices. The success of these models in finance is justified, on one hand, by their analitycal tractability (the marginal distribution can be determined through the characteristic function) and, on the other hand, by the ability on reproducing asymmetry and heavy tails in financial time series (see Cont and Tankov, 2003; Schoutens, 2003, for a general survey). A special attention deserves the process whose distribution at time one is a Normal Variance Mean Mixture. Particular cases widely applied in finance are the Variance Gamma process introduced by Madan and Seneta (1990), the Normal Inverse Gaussian (see Barndorff-Nielsen and Shephard, 2001), the Hyperbolic and the Generalized Hyperbolic (see Barndorff-Nielsen, 1977; Eberlein and Prause, 1998). Their main drawback is related to the independence of the increments that makes them inadequate in capturing the dynamic of higher moments (see Iacus, 2011, for formulas of some Lévy processes applied in option pricing).

A way to overcome these limits is by using stochastic volatility models for describing log return dynamics. There are two sources of risk in these models: the first drives the volatility dynamics and the second directly log returns. The main problem is that the volatility process is not observable in the market.

In discrete time the most commonly used class for modeling financial time series is the family of GARCH models (Engle, 1995). Despite the success in financial econometrics and risk management, their use for option pricing is not yet very well understood, as observed in Christoffersen et al. (2012). Monte Carlo technique is often used to compute option prices in GARCH models (see Duan, 1995; Duan and Simonato, 1998, for the efficiency of Monte Carlo estimator). Another approach is using approximate formulas based on Edgeworth expansion (see Duan et al., 1999, 2006). It is well known that the Monte Carlo procedure is time consuming when calibration exercise is considered, while the Edgeworth expansion seems to be less accurate for option pricing with long or medium time to maturity.

A major breakthrough occurred with the paper of Heston and Nandi (2000) where the authors derived a recursive procedure for the characteristic function of the log price at maturity, obtaining a semi analytical formula for European call options based on Inverse Fourier Transform, as in Carr and Madan (1999). Following the same idea a new class of GARCH models, namely affine GARCH, has been developed assuming different assumptions for the innovations. In particular, Christoffersen et al. (2006) considered the Inverse Gaussian innovations while Bellini and Mercuri (2007) Gamma innovations. Later Mercuri (2008) generalized the class of affine GARCH models assuming that log returns are conditionally Tempered Stable distributed (see Ornthanalai, 2008, for more details on affine GARCH models).

As observed in Christoffersen et al. (2006), the extreme asymmetry of the affine GARCH models with non-normal innovations gives an advantage for options with very short time to maturity. However, the fit is less accurate for options with medium maturity probably due to the fact that the medium time to maturity return distribution slowly converges to the Normal distribution.

To overcome this limit, starting from the affine GARCH model and assuming that the conditional distribution of log returns is a Normal Variance Mean Mixture, we construct a discrete time stochastic volatility model in a simple way. Indeed, substituting the mixing random variable with an affine GARCH, we obtain a recursive procedure for the computation of the characteristic function for the log-price at maturity. Option prices are obtained via Fourier transform. The introduction of this new class is motivated from the fact that affine models (usually in continuous time) are quite natural for option pricing but the discrete time models are easily estimated. Although the literature on affine stochastic volatility in continuous time is wide, the discrete counterpart did not receive the same attention. The substitution of the mixing r.v. with an affine GARCH process gives to our models the capability of capturing time dependence in financial times series, for instance persistence in squared returns. This affine GARCH process controls also the magnitude of the return movements and plays a similar role as the variance process in the continuous time models. Moreover, it generates time varying higher order moments. Volatility (see Chicago Board Options Exchange, 2003) and Skew (see Chicago Board Options Exchange, 2011) indexes cannot exist in a world with constant higher moments since they would be useless. Time-dependence of these moments is coherent with price movements observed in the market making our approach more realistic.

In our model, it is possible to extrapolate information from the VIX data and use it in option pricing. Indeed, we find a linear relation between the variance dynamics and the VIX^2. A similar result has been obtained in discrete time by Hao and Zhang (2013) under the GARCH assumption and, for these models, the procedure for extrapolating information from VIX in pricing Options on S&P500 has been considered recently in Kanniainen et al. (2014) while Liu et al. (2015) analyze how to assess the risk premium in GARCH(1,1), GJR, and HestonNandi models. However, our model is able to generate time-varying skewness and kurtosis that standard GARCH models can not reproduce.

The paper is organized as follows. Section 7.2 explains the construction of stochastic volatility models in discrete time. In Sect. 7.3 we prove that, in our setup, the VIX index is an autoregressive process with heteroskedastic innovations: we derive a linear relation between the unobservable variance and the current level of VIX index. In Sect. 7.4 we derive explicit formulas specifying the conditional distribution of log returns. In Sect. 7.5 empirical results using the implied volatility surface obtained from Bloomberg data provider are given. In Sect. 7.6 we draw some conclusions.

7.2 General Setup

In this section we propose a class of stochastic volatility models, in discrete time, through which we are able to price options using the information extrapolated from the VIX index.

Given a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$, we consider a market with two assets:

– riskless with dynamics: $B_t = B_{t-1}\exp(r)$
– risky with price dynamics:

$$\begin{aligned} S_t &= S_{t-1}\exp(X_t) \\ X_t &= r + \lambda_0 h_t + \lambda_1 V_t + \sigma\sqrt{V_t}Z_t \end{aligned} \tag{7.1}$$

where r is the deterministic free rate observed in the market; λ_0 and λ_1 are real valued model parameters while σ must be non negative; X_t is a discrete time stochastic process describing log returns; $Z_t \sim N(0,1),\ \forall t = 1,\ldots,T$ and is independent from V_t.

We require V_t to be an adapted positive process such that the conditional moment generating function (m.g.f. hereafter) of V_t given the information available at time $t-1$ is:

$$E[\exp(cV_t)|\mathcal{F}_{t-1}] = \exp(h_t f(c,\theta)) \tag{7.2}$$

and $\forall$ fixed vector θ, $\exists\,\delta > 0$ such that $\forall c \in (-\delta,\delta)$ the function $f(c,\theta) \in C^\infty$ and $f(0,\theta) = 0$. The vector θ contains the parameters of distribution V_t given information at time $t-1$. We assume h_t to be a predictable process defined as:

$$h_t = \alpha_0 + \alpha_1 V_{t-1} + \beta h_{t-1}. \tag{7.3}$$

The process h_t is positive if the parameters α_0, α_1 and β are non negative.

It is worth noting that if V_t is constant (i.e. $V_t = \bar{V},\ t = 1,2,\ldots$ and consequently $h_t = \bar{h}$), the sequence $\{X_t\}_{t=1,2,\ldots}$ is composed by i.i.d gaussian r.v.' s and a general sample path is centered in $r + \lambda_0\bar{h} + \lambda_1\bar{V}$. The magnitude of movements depends on the value of $\bar{V}$. In this case, the oscillating behaviour of returns in quiet and in turbulent markets can not be reproduced. The same observation holds if we assume the sequence $\{V_t\}_{t=1,2,\ldots}$ to be composed by i.i.d. random variables.

From (7.2) we have:

$$E[V_t|\mathcal{F}_{t-1}] = \left.\frac{\partial E[\exp(cV_t)|\mathcal{F}_{t-1}]}{\partial c}\right|_{c=0}.$$

Let $g(\theta)$ be defined as:

$$g(\theta) := \left.\frac{\partial f(c,\theta)}{\partial c}\right|_{c=0}, \tag{7.4}$$

the analytical expression for conditional mean of V_t becomes:

$$E[V_t|\mathcal{F}_{t-1}] = h_t g(\theta). \tag{7.5}$$

Adding and subtracting the quantity $\alpha_1 g(\theta)h_{t-1}$ in (7.3) we obtain for h_t a new representation:

$$h_t = \alpha_0 + (\alpha_1 g(\theta) + \beta)h_{t-1} + \alpha_1(V_{t-1} - g(\theta)h_{t-1}). \tag{7.6}$$

Observe that h_t is an AR(1) with heteroskedastic error $V_{t-1} - g(\theta)h_{t-1}$. Therefore, if we extrapolate from the market the realizations of h_t, the generalized least square technique gives us estimates for the quantities α_0, α_1, and $\alpha_1 g(\theta) + \beta$. In our model, the conditional variance evolves according to the stochastic process h_t:

$$Var\left[V_t\right|\mathcal{F}_{t-1}] = h_t \left.\frac{\partial^2 f(c,\theta)}{(\partial c)^2}\right|_{c=0}.$$

An essential requirement, based on empirical evidence, is the negative correlation between returns and volatility which implies:

$$Cov\left(V_t,X_t\right|\mathcal{F}_{t-1}) = \lambda_1 Var(V_t|\mathcal{F}_{t-1}) < 0, \tag{7.7}$$

meaning that λ_1 must be negative.

If we compute the conditional expectation of X_t we have:

$$E\left(X_t\left|\mathcal{F}_{t-1}\right.\right) = r + (\lambda_0 + \lambda_1 g\left(\theta\right))\, h_t. \tag{7.8}$$

Looking to relation in (7.8) is natural for a financial interpretation to require $\lambda_0 + \lambda_1 g\left(\theta\right) > 0$ since it implies a positive risk premium for the asset.

In the special case when $\sigma = 0$ the process describing X_t is an affine GARCH as in Christoffersen et al. (2006), Bellini and Mercuri (2007) and Mercuri (2008).

Our approach tries to generalize the Lévy processes built on the Normal Variance Mean Mixtures since we introduce a dependence structure. Indeed the conditional distribution evolves through time due to the predictable process h_t.

Both h_t and σ are crucial for the variability of the process X_t but σ does not introduce any heteroskedasticity in the model and for obtaining time dependent higher moments we need h_t to be defined as in (7.6). Through the predictable process h_t, we are able to generalize the Lévy process built on the Normal Variance Mean Mixture obtaining a distribuition of increments that evolves in time.

The next step is to show how to price a European call option with maturity T where the dynamics of the log returns for the risky asset is defined in (7.1). Here, we provide a simple recursive procedure through which we obtain the conditional m.g.f. of $\ln(S_T)$ using a similar approach as that introduced in Heston and Nandi (2000).

Proposition 1 *Under condition (7.2), the m.g.f. of the random variable* $\ln S_T$ *given the information at time t exists and is given by:*

$$E[\exp(c \ln(S_T))|\mathcal{F}_t] = S_t^c \exp[A(t;T,c) + B(t;T,c)h_{t+1}].$$

The time-dependent coefficients $A(t;T,c)$ *and* $B(t;T,c)$ *are:*

$$\begin{cases} A(t;T,c) = cr + A(t+1;T,c) + \alpha_0 B(t+1;T,c) \\ B(t;T,c) = c\lambda_0 + \beta B(t+1;T,c) + \\ \qquad\qquad f(c\lambda_1 + \alpha_1 B(t+1;T,c) + \frac{c^2\sigma^2}{2}, \theta) \end{cases} \tag{7.9}$$

with the following final conditions:

$$\begin{aligned} A(T;T,c) &= 0 \\ B(T;T,c) &= 0. \end{aligned}$$

(see Appendix A.1)
The existence of m.g.f. allows us to obtain the characteristic function and the distribution function is achieved by the inverse Fourier transform.

Our aim is to price options and compute implied volatility indexes. In order to ensure the martingale condition under Q measure, we use the following proposition.

Proposition 2 *Under the assumptions* $E(S_t) < +\infty$ *and* $\lambda_0 = -f(\lambda_1 + \frac{\sigma^2}{2}, \theta)$*, the discounted price is a martingale.*

(see Appendix A.2)

We have obtained in Proposition 1 the m.g.f. for the underlying of a call option. The next step is the evaluation of a European call option as in Heston (1993)

$$\begin{aligned} C(K,T) &= S_0\Pi_1 - Ke^{-rT}\Pi_2 \\ \Pi_1 &= \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} \Re\left(\frac{K^{-iu}E_0^Q[S_T^{i(u-i)}]}{iuE_0^Q[S_T]}\right) du \\ \Pi_2 &= \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} \Re\left(\frac{K^{-iu}E_0^Q[S_T^{iu}]}{iu}\right) du \end{aligned}$$

The exercise probabilities Π_1 and Π_2 can be computed following Feller (1968).

7.3 VIX Index

In this Section we show how to derive the Volatility Implied Index (VIX) in our model. In particular, the linear relation between VIX and the process h_t is derived. From a theoretical point of view, this relation implies that the VIX is a mean-reverting autoregressive process with heteroskedastic errors. A similar result has been proposed in Zhang and Zhu (2006) under the assumption that the SPX dynamics is described by Heston (1993). The methodology for computing the VIX index is based on the replication of a variance swap (see Demeterfi et al., 1999) and the current level of VIX is related to the value of the portfolio composed by out-of-the money call/put options on the S&P500. Assuming that the strike prices vary continuously from 0 to $+\infty$, the VIX squared formula is the following:

$$\left(\frac{VIX_t}{100}\right)^2 = \frac{2e^{r(T-t)}}{T-t}\left[\int_0^{S^*}\frac{1}{K^2}P(S_t,K)dK + \int_{S^*}^{+\infty}\frac{1}{K^2}C(S_t,K)dK\right]$$
$$= \frac{2e^{r(T-t)}}{T-t}\left[E_t^Q\left(\frac{S_T-S^*}{S^*} - \ln\left(\frac{S_T}{S^*}\right)\right)\right]. \tag{7.10}$$

$C(S_t,K)$ and $P(S_t,K)$ are out-of-the money call and put option prices. S^* is the forward price of the SPX index.

The main result of our model is reported in the following proposition.

Proposition 3 *Under the conditions:*

$$\begin{aligned} &\alpha_1 g(\theta) + \beta < 1 \\ &\lambda_1 g(\theta) - f\left(\lambda_1 + \tfrac{\sigma^2}{2}, \theta\right) \le 0 \\ &h_{t+1} > 0 \end{aligned} \tag{7.11}$$

the VIX squared is an affine linear function of the predictable process h_t:

$$\left(\frac{VIX_t}{100}\right)^2 = -\frac{2e^{r(T-t)}}{T-t}\left[C(t;T) + D(t;T)h_{t+1}\right] \tag{7.12}$$

where $C(t;T)$ and $D(t;T)$ are functions of the model parameters, given by:

$$\begin{cases} C(t;T) = \alpha_0\left[\lambda_1 g(\theta) + \lambda_0\right]\left\{\frac{T-t-1-[\alpha_1 g(\theta)+\beta]\frac{1-[\alpha_1 g(\theta)+\beta]^{(T-t)-1}}{1-[\alpha_1 g(\theta)+\beta]}}{1-[\alpha_1 g(\theta)+\beta]}\right\} \\ D(t;T) = \left[\lambda_1 g(\theta) + \lambda_0\right]\frac{1-[\alpha_1 g(\theta)+\beta]^{T-t}}{1-[\alpha_1 g(\theta)+\beta]} \end{cases} \tag{7.13}$$

with $T-t = 30$ days.

(See Appendix A.3)

Considering the fact that VIX is a measure of the 30 days implied volatility on S&P500, Eq. (7.12) becomes:

$$\left(\frac{VIX_t}{100}\right)^2 = -\frac{2e^{r30}}{30}\left[C_{30} + D_{30}h_{t+1}\right]$$

where r is the one month Libor rate on daily basis.

We define the adjusted VIX as:

$$VIX_t^{adj} = -\frac{30}{2e^{r30}}\frac{VIX_t^2}{10^4}.$$

Notice that $VIX_t^{adj} < 0 \ \forall t$ since it is a decreasing linear transformation of the VIX squared.

Using Proposition 3 we have:

$$VIX_t^{adj} = C_{30} + D_{30}h_{t+1} \ \Rightarrow \ h_{t+1} = \frac{VIX_t^{adj} - C_{30}}{D_{30}}. \tag{7.14}$$

The requirement $h_{t+1} > 0$ implies that $0 > VIX_t^{adj} > C_{30} \ \forall t$.

Using the definition (7.6) of h_t, we have following proposition:

Proposition 4 *Under the same conditions of Proposition 3, defining the heteroskedastic error term* $\tau_t := \alpha_1(V_t - g(\theta)h_t)D_{30}$*, the* VIX_t^{adj} *is an* $AR(1)$ *defined as:*

$$VIX_t^{adj} = int + slope VIX_{t-1}^{adj} + \tau_t$$

where

$$\begin{cases} int = 30\alpha_0\left(\lambda_1 g(\theta) + \lambda_0\right) \\ slope = \alpha_1 g(\theta) + \beta \end{cases}$$

(see Appendix A.4)

Given the model parameters, the current and the one-day-ahead VIX level we have the heteroskedastic error term defined as:

$$\tau_{t+1} = VIX_{t+1} - int - slope VIX_t.$$

From Eq. (7.14) we extract h_{t+1} and obtain the value of the main "unobservable" variable of our model, i.e. V_{t+1}:

$$V_{t+1} = g(\theta) + \frac{\tau_{t+1}}{\alpha_1 D_{30}}.$$

Once estimated *int* and *slope* we can redefine D_{30} and C_{30} in order to extrapolate a multiple of h_{t+1} from the quoted VIX_t. In particular we get:

$$D_{30} = \frac{D^*_{30}}{\alpha_0} = \frac{int\left(1 - slope^{30}\right)}{30 * (1 - slope)} \frac{1}{\alpha_0}$$

$$C_{30} = \left[\frac{29 - slope\frac{1-slope^{29}}{1-slope}}{1 - slope} \right] \frac{int}{30}$$

$$\frac{VIX^{adj}_t - C_{30}}{D^*_{30}} = \frac{h_{t+1}}{\alpha_0} > 0$$

The quantity $\frac{h_{t+1}}{\alpha_0}$ can be used to compute the m.g.f. of $\ln(S_T)|\mathcal{F}_t$ needed in option pricing.

If $slope < 1$, VIX^{adj}_t is mean reverting. The long term mean and the reverting speed are respectively:

$$\frac{int}{1 - slope}, \quad 1 - slope.$$

The conditional mean of the error term is zero but we are in presence of heteroskedasticity:

$$E\left[\tau_t | \mathcal{F}_{t-1}\right] = 0, \quad Var\left[\tau_t | \mathcal{F}_{t-1}\right] = \alpha_1^2 D_{30}^2 Var\left[V_t | \mathcal{F}_{t-1}\right].$$

Although $Cov\left[\tau_{t+1}, \tau_t | \mathcal{F}_{t-1}\right] = 0$ and $Cov\left[\tau_{t+1}, \tau_t^2 \middle| \mathcal{F}_{t-1}\right] = 0$, the error time-dependence structure is more complex than a linear one. The following quantities are different from zero and time dependent:

$$Cov\left[\tau^2_{t+1}, \tau_t | \mathcal{F}_{t-1}\right] = \alpha_1^3 D_{30}^3 \left.\frac{\partial^2 f}{(\partial c)^2}\right|_{c=0} \left[\alpha_0 + \alpha_1 \left.\frac{\partial^2 f}{(\partial c)^2}\right|_{c=0}\right] h_t$$

$$Cov\left[\tau^2_{t+1}, \tau^2_t | \mathcal{F}_{t-1}\right] = \alpha_1^4 D_{30} \left.\frac{\partial^2 f(c,\theta)}{(\partial c)^2}\right|_{c=0} \left[\alpha_0 + (\alpha_1 g(\theta) + \beta) \left.\frac{\partial^2 f(c,\theta)}{(\partial c)^2}\right|_{c=0} h_t^2 + \alpha_1^2 \mu_3\right]$$

where $\mu_3 = E\left[(V_t - g(\theta))^3 \middle| \mathcal{F}_{t-1}\right]$.

7.4 Special Cases

The conditional distribution of log returns belongs to the family of Normal Variance Mean Mixture since Z_t in (7.1) is normally distributed. A univariate Normal Variance Mean Mixture (see Barndorff-Nielsen et al., 1982) is a random variable defined as:

$$X \stackrel{d}{=} \mu + \lambda V + \sigma \sqrt{V} Z$$

where Z and V are independent univariate random variables, $Z \sim N(0, 1)$, and V is defined on the positive real line. Below we introduce three special cases of our approach where the conditional distribution of log returns is respectively Variance Gamma (see Madan and Seneta, 1990), Normal Inverse Gaussian (see Barndorff-Nielsen and Shephard, 2001) and Normal Tempered Stable (see Barndorff-Nielsen and Shephard, 2001).

7.4.1 Dynamic Variance Gamma

Assuming that the affine GARCH process V_t is conditionally Gamma distributed (see Bellini and Mercuri, 2007) then X_t in (7.1) follows a Dynamic Variance Gamma model introduced by Bellini and Mercuri (2011).

The conditional m.g.f. of the V_t is:

$$\begin{array}{ll} E\left[e^{cV_t} \middle| \mathcal{F}_{t-1} \right] & = \exp\left[-h_t \ln\left(1-c\right) \right] \\ f(c, \theta) & = -\ln\left(1-c\right) \\ g(\theta) & = 1. \end{array}$$

System (7.9) becomes:

$$\begin{cases} A(t; T, c) = cr + A(t+1; T, c) + \alpha_0 B(t+1; T, c) \\ B(t; T, c) = c\lambda_0 + \beta B(t+1; T, c) + \\ \qquad\qquad - \ln\left(1 - c\lambda_1 - \alpha_1 B(t+1; T, c) - \frac{c^2\sigma^2}{2}\right). \end{cases} \tag{7.15}$$

System (7.13) becomes:

$$\begin{cases} C(t; T, c) = \alpha_0 \left(\lambda_1 + \lambda_0\right) \left\{ \frac{(T-t) - (\alpha_1+\beta) \frac{1-(\alpha_1+\beta)^{T-t-1}}{1-(\alpha_1+\beta)}}{1-(\alpha_1+\beta)} \right\} \\ D(t; T, c) = \left(\lambda_1 + \lambda_0\right) \frac{1-(\alpha_1+\beta)^{T-t}}{1-(\alpha_1+\beta)}. \end{cases} \tag{7.16}$$

with final conditions $C(T; T, c) = 0$ and $D(T; T, c) = 0$. We have the following restrictions on the parameters:

$$\begin{cases} \lambda_1 \le 0 \\ \lambda_0 = \ln\left(1 - \lambda_1 - \frac{\sigma^2}{2}\right) \\ \alpha_1 + \beta \le 1 \\ \lambda_1 + \ln\left(1 - \lambda_1 - \frac{\sigma^2}{2}\right) \le 0. \end{cases} \tag{7.17}$$

7.4.2 Dynamic Normal Inverse Gaussian

If the affine GARCH process V_t is conditionally Inverse Gaussian distributed (see Christoffersen et al., 2006) than log-returns X_t, given the information at time $t-1$, have a Normal Inverse Gaussian distribution (see Barndorff-Nielsen, 1997).

The density of a Inverse Gaussian distribution is:

$$f_V(v) = \frac{h_t}{\sqrt{2\pi v^3}} \exp\left[-\frac{1}{2}\left(\sqrt{v} - \frac{h_t}{\sqrt{x}}\right)^2\right].$$

The conditional m.g.f. of the V_t is:

$$\begin{aligned} E\left[e^{cV_t}\middle|\mathcal{F}_{t-1}\right] &= \exp\left[h_t\left(1-\sqrt{1-2c}\right)\right] \\ f(c,\theta) &= \left(1-\sqrt{1-2c}\right) \\ g(\theta) &= 1. \end{aligned}$$

System (7.9) becomes:

$$\begin{cases} A(t;T,c) = xr + A(t+1;T,c) + \alpha_0 B(t+1;T,c) \\ B(t;T,c) = c\lambda_0 + \beta B(t+1;T,c) + \\ \qquad\qquad \sqrt{1-2\left(c\lambda_1 + \alpha_1 B(t+1;T,c) + \frac{c^2\sigma^2}{2}\right)}. \end{cases} \tag{7.18}$$

System (7.13) becomes

$$\begin{cases} C(t;T,c) = \alpha_0\left(\lambda_1+\lambda_0\right)\left\{\dfrac{(T-t)-(\alpha_1+\beta)\frac{1-(\alpha_1+\beta)^{T-t-1}}{1-(\alpha_1+\beta)}}{1-(\alpha_1+\beta)}\right\} \\ D(t;T,c) = (\lambda_1+\lambda_0)\frac{1-(\alpha_1+\beta)^{T-t}}{1-(\alpha_1+\beta)} \end{cases} \tag{7.19}$$

with final conditions $C(T;T,c) = 0$ and $D(T;T,c) = 0$. We have the following restrictions on the parameters:

$$\begin{cases} \lambda_1 \le 0 \\ \lambda_0 = -\left(1-\sqrt{1-2\left(\lambda_1+\frac{\sigma^2}{2}\right)}\right) \\ \lambda_1 - 1 + \sqrt{1-2\left(\lambda_1+\frac{\sigma^2}{2}\right)} < 0 \\ \alpha_1+\beta < 0 \end{cases} \tag{7.20}$$

7.4.3 Dynamic Normal Tempered Stable

Consider the affine process V_t proposed in Mercuri (2008) then log returns follow a conditional Normal Tempered Stable as introduced in Barndorff-Nielsen and Shephard (2001). We recall that the Normal Tempered Stable is obtained as a Normal Variance Mean Mixture where the mixing density is a Tempered Stable (see Tweedie, 1984) that is obtained by tempering the tail of a positively skewed $\alpha-$Stable distribution. The Normal Tempered Stable has as special cases the Variance Gamma and the Normal Inverse Gaussian.

The conditional m.g.f. of $V_t|\mathcal{F}_{t-1}$ is:

$$E\left[e^{cV_t}\middle|\,\mathcal{F}_{t-1}\right] = \exp\left[h_t b\left(1-(1-2cb^{-1/\alpha})^{\alpha}\right)\right] \tag{7.21}$$

where $\alpha \in (0, 1)$ and $b > 0$.

Comparing (7.21) with (7.2), we have:

$$f(c,\theta) = b\left(1-(1-2cb^{-1/\alpha})^{\alpha}\right)$$

and

$$g(\theta) = 2\alpha b^{(\alpha-1)/\alpha}.$$

Applying Proposition 1, we obtain the recursive system of equations for time dependent coefficients:

$$\begin{cases} A(t;T,c) = cr + A(t+1;T,c) + \alpha_0 B(t+1;T,c) \\ B(t;T,c) = c\lambda_0 + \beta B(t+1;T,c) + \\ \qquad\qquad b\left\{1-\left[1-2b^{-\frac{1}{\alpha}}\left(c\lambda_1 + \alpha B(t+1;T,c) + \frac{c^2\sigma^2}{2}\right)\right]^{\alpha}\right\}. \end{cases} \tag{7.22}$$

From Proposition 2 we have the following constraint:

$$\lambda_0 = -b\left[1-\left(1-2\left(\lambda_1+\frac{\sigma^2}{2}\right)b^{1/\alpha}\right)^{\alpha}\right]$$

and, implementing the Fast Fourier Transform, we price the European call option.

Using Proposition 3, we obtain the following time varying coefficients that allow us to extrapolate h_t from current level of VIX:

$$\begin{cases} C(t;T) = \alpha_0\left(2\alpha b^{(\alpha-1)/\alpha}\lambda_1 + \lambda_0\right) * \\ \qquad\qquad * \left\{\dfrac{(T-t)-(2\alpha b^{(\alpha-1)/\alpha}\alpha_1+\beta)\frac{1-\left(2\alpha b^{(\alpha-1)/\alpha}\alpha_1+\beta\right)^{T-t-1}}{1-\left(2\alpha b^{(\alpha-1)/\alpha}\alpha_1+\beta\right)}}{1-\left(2\alpha b^{(\alpha-1)/\alpha}\alpha_1+\beta\right)}\right\} \\ D(t;T) = (2\alpha b^{(\alpha-1)/\alpha}\lambda_1 + \lambda_0)\dfrac{1-\left(2\alpha b^{(\alpha-1)/\alpha}\alpha_1+\beta\right)^{T-t}}{1-\left(2\alpha b^{(\alpha-1)/\alpha}\alpha_1+\beta\right)} \end{cases}. \tag{7.23}$$

In this case the condition (7.11) becomes:

$$\begin{cases} 2\alpha b^{(\alpha-1)/\alpha}\lambda_1 - b\left[1-\left(1-2\left(\lambda_1+\frac{\sigma^2}{2}\right)b^{-1/\alpha}\right)^{\alpha}\right] \leq 0. \\ \alpha_1+\beta<1 \end{cases} \tag{7.24}$$

7.5 Empirical Analysis

We investigate in details the ability of our models to reproduce the behavior of European option prices on SPX index. We have two main objectives: replicate the market option volatilities and compare the theoretical VIX derived in our models with the observed one. The dataset is composed by the implied volatility surfaces observed each Wednesday going from May 2011 to April 2012, moneyness ranging from 0.9 to 1.1 and time to maturity 30, 60 and 90 days (the total number of observations is 1008). We choose Wednesday's observations to remove possible weekend effects as those discussed in French (1980). From Eq. (1) we see that we need the term structure of the risk-free rate in order to compute the m.g.f of the variable $\ln S_T$. The Libor curve can be a possible choice though we know it is not the only one. We downloaded the needed curve from Bloomberg.

The first Wednesdays of each month are the in-sample data (231 observations). The remaining dataset (777 observations) is used for the out-of-sample analysis. We calibrate the model in each in-sample period. The values obtained for the parameters are used as input for the out-of-sample analysis. The error measure considered is:

$$\sqrt{percMSE} = \sqrt{\frac{\sum_{k=1}^{K}\sum_{t=1}^{T}\left[\frac{\sigma^{mkt}(k,t)-\sigma^{theo}(k,t)}{\sigma^{mkt}(k,t)}\right]^2}{N_T * N_K}}$$

where $\sigma^{mkt}(k,t)$, $\sigma^{theo}(k,t)$ are respectively the implied volatilities observed in the market and those obtained by the models. N_T, N_K represent respectively the number of the available maturities and strikes.

Tables 7.1, 7.2 and 7.3 report the values of the calibrated parameters and the corresponding in-sample errors.

Our calibration exercise takes into account the possibility of extrapolating the latent process h_t directly from the VIX index. We find that for the DNTS model the in-sample errors are the lowest except only in one case where the DNIG model has the best performance. This result strongly supports our initial guess that two additional parameters would allow to better capture the market dynamics. Observe that if $b = 2^a$ and $\alpha = \frac{1}{a}$ for $a \to 0$ we obtain the DVG model, while for $b = 1$ and $\alpha = \frac{1}{2}$ the model is the DNIG.

The out-of-sample results suggest the use of the DNTS model in the considered dataset. Indeed, computing the $\sqrt{percMSE}$ on the entire out-of-sample data, we find that the DNTS reaches an error level of 5.05% which is a reduction error of 21.10%

Table 7.1 Calibrated parameters for the DVG model in the in-sample period

In sample calibration for DVG							
Date	λ_0	λ_1	σ	α_0	α_1	β	Perc. error
04-May-2011	0.012	−0.012	0.014	0.033	0.493	0.379	0.048
01-Jun-2011	0.036	−0.039	0.069	0.009	0.274	0.148	0.084
06-Jul-2011	0.005	−0.005	0.006	0.033	0.344	0.633	0.027
03-Aug-2011	0.034	−0.035	0.001	0.060	0.317	0.000	0.037
07-Sep-2011	0.008	−0.008	0.011	0.032	0.538	0.444	0.015
05-Oct-2011	0.051	−0.053	0.029	0.028	0.155	0.484	0.024
02-Nov-2011	0.095	−0.100	0.007	0.018	0.057	0.085	0.039
07-Dec-2011	0.060	−0.062	0.007	0.024	0.008	0.454	0.052
04-Jan-2012	0.019	−0.019	0.020	0.023	0.207	0.644	0.048
01-Feb-2012	0.036	−0.038	0.056	0.017	0.014	0.157	0.048
07-Mar-2012	0.042	−0.043	0.029	0.000	0.000	1.000	0.088

Table 7.2 Calibrated parameters for the DNIG model in the in-sample period

In sample calibration for DNIG							
Date	λ_0	λ_1	σ	α_0	α_1	β	Perc. error
04-May-2011	0.049	−0.052	0.062	0.006	0.012	0.572	0.039
01-Jun-2011	0.047	−0.050	0.061	0.006	0.016	0.604	0.029
06-Jul-2011	0.009	−0.009	0.011	0.009	0.168	0.816	0.024
03-Aug-2011	0.035	−0.036	0.042	0.029	0.212	0.059	0.022
07-Sep-2011	0.067	−0.072	0.075	0.017	0.120	0.113	0.022
05-Oct-2011	0.007	−0.008	0.010	0.028	0.427	0.564	0.007
02-Nov-2011	0.060	−0.064	0.066	0.007	0.081	0.674	0.019
07-Dec-2011	0.046	−0.048	0.057	0.005	0.008	0.867	0.024
04-Jan-2012	0.029	−0.030	0.019	0.019	0.065	0.733	0.057
01-Feb-2012	0.030	−0.031	0.042	0.023	0.269	0.109	0.034
07-Mar-2012	0.013	−0.014	0.015	0.010	0.211	0.760	0.026

with respect to DNIG (the second best model). To deeply analyse the out of sample error, Fig. 7.1 reports the results obtained in 36 out-of-sample Wednesdays. In 72% of the cases the DNTS shows a lower error level than the other two while the DNIG has the lowest error level only in 14% of the cases.

We remark that in our model the square of the VIX is an autoregressive process. The conditional expected value of the VIX is not available in a closed form formula. However, using Jensen's inequality, we easily derive the following upper bound that we use in our analysis:

$$E\left[VIX_{t+1}\middle|\,\mathcal{F}_t\right] = E\left[\sqrt{VIX_{t+1}^2}\,\middle|\,\mathcal{F}_t\right] \le \sqrt{E\left[VIX_{t+1}^2\middle|\,\mathcal{F}_t\right]} = VIX_{t+1}^{ub}.$$

Table 7.3 Calibrated parameters for the DNTS model in the in-sample period

In sample calibration for DNTS									
Date	λ_0	λ_1	σ	α_0	α_1	β	b	a	Perc. error
04-May-2011	0.212	−0.107	0.013	0.002	0.363	0.276	0.814	0.990	0.009
01-Jun-2011	0.066	−0.042	0.039	0.011	0.296	0.000	0.800	0.750	0.019
06-Jul-2011	0.005	−0.005	0.006	0.039	0.380	0.596	1.000	0.500	0.025
03-Aug-2011	0.052	−0.027	0.008	0.012	0.510	0.000	0.897	0.955	0.007
07-Sep-2011	0.005	−0.006	0.009	0.051	0.658	0.409	0.962	0.413	0.015
05-Oct-2011	0.012	−0.016	0.026	0.001	0.116	0.910	0.946	0.345	0.004
02-Nov-2011	0.003	−0.003	0.006	0.133	0.806	0.233	0.863	0.351	0.011
07-Dec-2011	0.085	−0.043	0.010	0.004	0.473	0.072	0.854	0.975	0.005
04-Jan-2012	0.003	−0.005	0.007	0.105	0.772	0.449	1.000	0.341	0.018
01-Feb-2012	0.100	−0.053	0.021	0.001	0.104	0.803	0.800	0.934	0.014
07-Mar-2012	0.018	−0.011	0.007	0.032	0.539	0.090	0.965	0.803	0.024

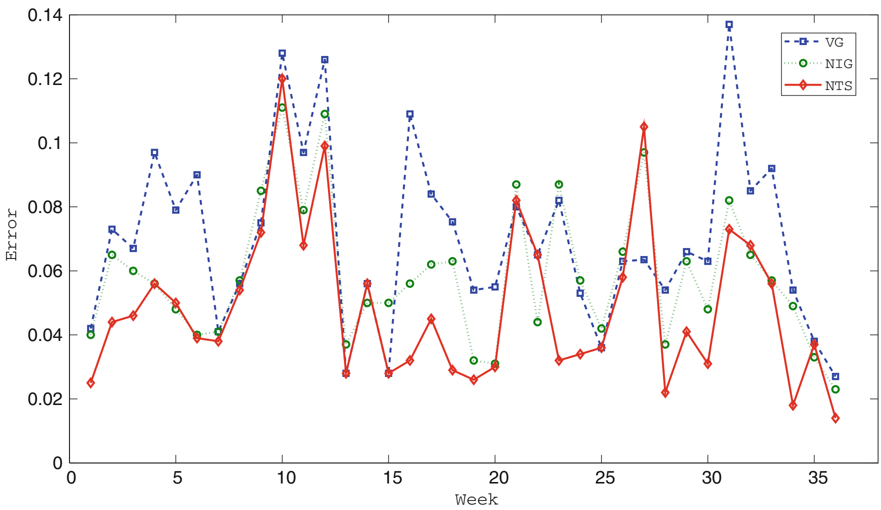


Fig. 7.1 Out of sample weekly comparison

Using Proposition 4 and Eq. (7.14), our upper bound becomes:

$$VIX_{t+1}^{ub} = \sqrt{-\frac{2e^{30r}10^4}{30} int + slope VIX_t^2}$$

where all quantities are on daily basis and the year conversion is necessary for comparison with its observed level.

We calibrate the model on the first Wednesday of each month (in total there are 12 calibration periods). The resulting parameters are maintained fixed until the next

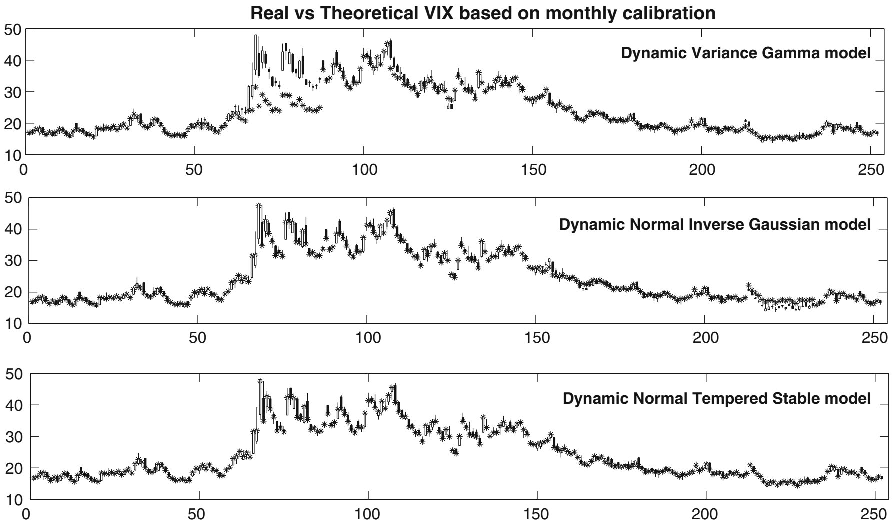


Fig. 7.2 Comparison between the predict VIX (*upper bound* $*$) and next day open, closed, min, max VIX level using monthly calibration

in-sample day. From Figs. 7.2, 7.4 and Table 7.5 we observe that the DVG model displays the worst performance.

Instead of having fixed parameters for the entire month we can decide to make the recalibration period dynamic. Intuitively, if the market conditions change a lot (i.e. we observe a jump of the implied volatility), it is reasonable to think that in order to have a better prediction for the VIX level we must update the model parameters. This update for us means to recalibrate the model using the option volatilities observed after the jump has occurred.

We face the problem of defining the jump in terms of relative daily variation of the VIX Index level. If the observed VIX level is lower than 30% we recalibrate if the next day relative variation is higher than 30%. For example if the current level of VIX is 15% we recalibrate the model if the next day value is higher than 20% or lower than 10%. For higher levels of the VIX index (more than 30%) the required daily relative variation is fixed at 25%. This decision comes from the fact that VIX levels higher than 39% are rarely observed. In Fig. 7.3 we report a comparison between the VIX and S&P500 for the considered dates.

The number of calibrations is reduced from 12 (when the parameters for the entire month are fixed) to 9 (when decision is dependent on the VIX level). Comparing the results reported in Tables 7.4 and 7.5, the error term, defined as $\sqrt{E\left[(VIX_{mkt} - VIX_{ub})^2\right]}$, is reduced when the calibration time is based on VIX index level. This is also confirmed from Figs. 7.3 and 7.4.

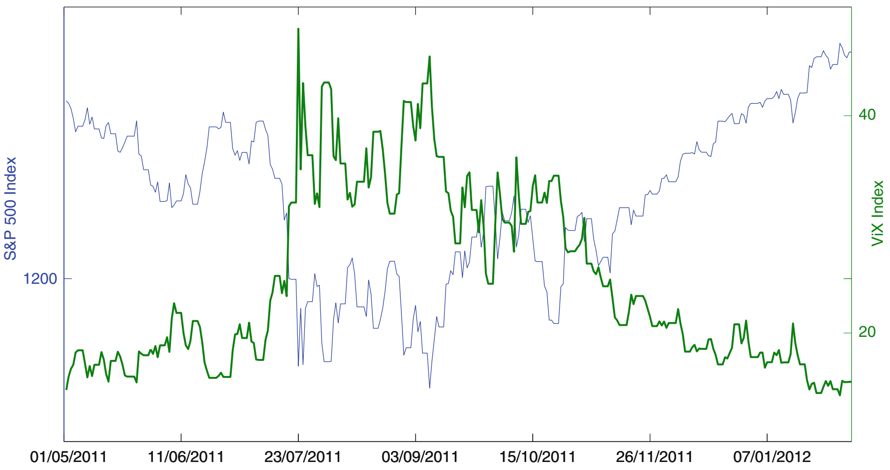


Fig. 7.3 Comparison between VIX and S&P500

Table 7.4 Errors obtained when the calibration is done the first Wednesday of each month

	DVG (%)	DNIG (%)	DNTS (%)
Open	1,111	0,029	0,140
Closing	0,967	0,173	0,004
High	2,187	1,047	1,216
Low	0,028	1,167	0,999

Table 7.5 Errors obtained when the calibration decision depends on the VIX index level

	DVG (%)	DNIG (%)	DNTS (%)
Open	0,589	0,005	0,080
Closing	0,445	0,139	0,064
High	1,665	1,081	1,156
Low	0,550	1,133	1,059

The choice of the DNTS showed in the calibration exercise seems to be weaker when we try to forecast the VIX index level. In particular, the DNIG seems to behave better in some extreme market conditions.

7.6 Conclusions

In this paper we proposed a class of discrete time stochastic volatility models. We started from the affine GARCH model and assumed that the conditional distribution of log returns is a Normal Variance Mean Mixture with support the entire real line. We obtained a recursive procedure for the computation of the characteristic function for the log-price at maturity. Option prices were than obtained via Fourier transform.

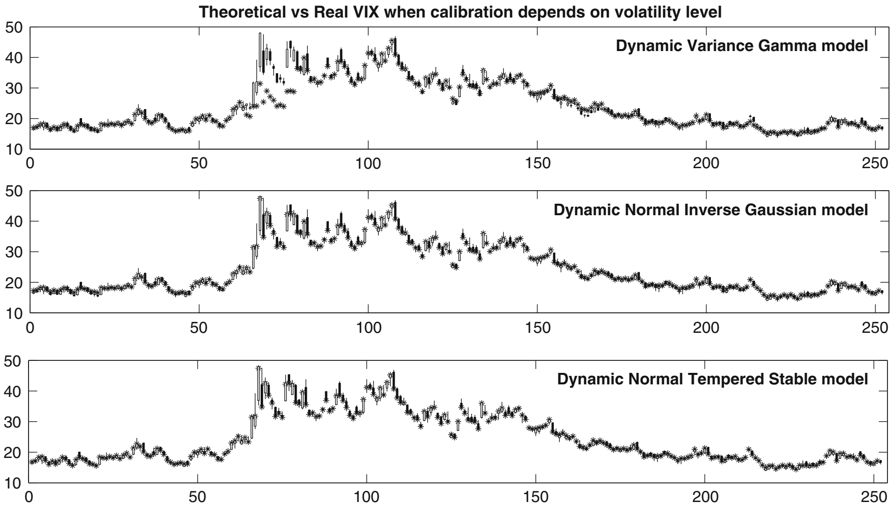


Fig. 7.4 Comparison between the predict VIX (*upper bound* $*$) and next day open, closed, min, max VIX level using monthly calibration

In our model, it is possible to extrapolate information from the VIX data. The VIX^2 index resulted to be an autoregressive process and the information extracted was used for pricing options on S&P500.

We specified some special cases for our general model. The Dynamic Normal Inverse Gaussian based model resulted to be more flexible in capturing market dynamics especially in turbulent periods.

A Appendix

A.1 Conditional Moment Generating Function

Following the approach proposed in Heston and Nandi (2000) we derive a system of recursive equations for the time dependent coefficients of the conditional m.g.f. of the random variable $\ln(S_T)$ given the available information at time t. We want to prove that the conditional m.g.f. is given by the following formula:

$$E_t\left[\exp\left(c\ln\left(S_T\right)\right)\middle|\,\mathcal{F}_t\right] = S_t^c \exp\left[A\left(t;T,c\right) + B\left(t;T,c\right)h_{t+1}\right]. \tag{A.1}$$

We use the mathematical induction method.

1. We observe that relation (A.1) holds at time T since $A(T;T,c) = 0$ and $B(T;T,c) = 0$.

2. We suppose the relation (A.1) holds at time $t+1$ and, by the law of iterated expectations, we prove it at time t.

$$\begin{aligned}
&E\left[E\left[S_T^c\middle|\mathcal{F}_{t+1}\right]\middle|\mathcal{F}_t\right] = E\left[\exp\left[A\left(t+1;T,c\right)+B\left(t+1;T,c\right)h_{t+2}\right]\middle|\mathcal{F}_t\right]\\
&= E\left[\exp\left[c\ln\left(S_T\right)+cr+A(t+1;T,c)\right.\right.\\
&+\ c\lambda_0 h_{t+1}+c\lambda_1 V_{t+1}+c\sigma\sqrt{V_{t+1}}Z_{t+1}+\\
&\left.\left.+\alpha_0 B\left(t+1;T,c\right)+\alpha_1 B\left(t+1;T,c\right)V_{t+1}+\beta B\left(t+1;T,c\right)h_{t+1}\ \right]\middle|\mathcal{F}_t\right]\\
&= S_t^c\exp\left[cr+A\left(t+1;T,c\right)+\alpha_0 B\left(t+1;T,c\right)+\left(c\lambda_0+\beta B\left(t+1;T,c\right)\right)h_{t+1}\right]*\\
&*E\left[\exp\left[\left(c\lambda_1+\alpha_1 B\left(t+1;T,c\right)+\frac{c^2\sigma^2}{2}\right)V_{t+1}\right]\middle|\mathcal{F}_t\right].
\end{aligned} \tag{A.2}$$

Using the conditional m.g.f. of the r.v. V_{t+1}, Eq. (A.2) becomes:

$$\begin{aligned}
E\left[E\left[S_T^c\middle|\mathcal{F}_{t+1}\right]\middle|\mathcal{F}_t\right] = {}& S_t^c\exp\left[cr+A\left(t+1;T,c\right)+\alpha_0 B\left(t+1;T,c\right)+\right.\\
&\left.+\left(c\lambda_0+\beta B\left(t+1;T,c\right)+f\left(c\lambda_1+\alpha_1 B\left(t+1;T,c\right)+\frac{c^2\sigma^2}{2},\theta\right)\right)h_{t+1}\right]
\end{aligned} \tag{A.3}$$

By comparing the expression obtained in Eq. (A.3) with (A.1) we obtain the following recursive system:

$$\begin{cases}
A(t;T,c) = cr + A(t+1;T,c) + \alpha_0 B(t+1;T,c)\\
B(t;T,c) = c\lambda_0 + \beta B(t+1;T,c) +\\
\qquad\qquad f(c\lambda_1 + \alpha_1 B(t+1;T,c) + \frac{c^2\sigma^2}{2},\theta)
\end{cases} \tag{A.4}$$

with $A(T;T,c) = 0$ and $B(T;T,c) = 0$.

A.2 Martingale Condition

We want to prove that $\forall s \le t$:

$$\lambda_0 = -f(\lambda_1 + \frac{\sigma^2}{2};\theta) \overset{(1)}{\Longrightarrow} E\left[\frac{S_t}{e^r}\middle|\mathcal{F}_{t-1}\right] = S_{t-1} \overset{(2)}{\Longrightarrow} E\left[\frac{S_t}{e^{r(t-s)}}\middle|\mathcal{F}_s\right] = S_s. \tag{A.5}$$

$(\overset{(1)}{\Longrightarrow})$

We assume r to be constant but the proof holds even assuming r to be a predictable process. By simple calculus, we obtain:

$$E\left[\frac{S_t}{e^r}\middle|\mathcal{F}_{t-1}\right] = S_{t-1}\exp\left[\left(\lambda_0 + f\left(\lambda_1+\frac{\sigma^2}{2};\theta\right)\right)h_{t-1}\right] \tag{A.6}$$

substituting $\lambda_0 = -f(\lambda_1 + \frac{\sigma^2}{2};\theta)$ in (A.6) we obtain the result.

$(\overset{(2)}{\Longrightarrow})$

By the iterated law of conditional expectation we have:

$$E\left[\frac{S_t}{e^{r(t-s)}}\middle|\mathcal{F}_s\right] = E\left[E\left[\frac{S_t}{e^{r(t-s)}}\middle|\mathcal{F}_{t-1}\right]\middle|\mathcal{F}_s\right]$$

$$= E\left[\frac{1}{e^{r(t-s-1)}}\underbrace{E\left[\frac{S_t}{e^r}\middle|\mathcal{F}_{t-1}\right]}_{S_{t-1}}\middle|\mathcal{F}_s\right]$$

$$= \ldots = E\left[\frac{S_{s+1}}{e^r}\middle|\mathcal{F}_s\right] = S_s.$$

A.3 VIX Index: Derivation Formula

We derive an analytical formula for the VIX index when the dynamics of S&P 500 belongs to our class. Defined S^* as the forward price of S_t with maturity $T-t$, we start from the VIX definition:

$$\left(\frac{VIX_t}{100}\right)^2 = \frac{2e^{r(T-t)}}{T-t}\left[\underbrace{E^Q\left[\frac{S_T-S^*}{S^*}\middle|\mathcal{F}_t\right]}_{(*)} - \underbrace{E^Q\left[\ln\left(\frac{S_T}{S^*}\right)\middle|\mathcal{F}_t\right]}_{(**)}\right].$$

The quantity in $(*)$ is 0 since:

$$E^Q\left[\frac{S_T-S^*}{S^*}\middle|\mathcal{F}_t\right] = \frac{1}{S_t e^{r(T-t)}}E^Q\left[S_T|\,\mathcal{F}_t\right] - 1 = 0.$$

Given the spot price S_t, we have $S_T = S_t \exp\left(\sum_{d=t+1}^T X_d\right)$ and by substituting in $(**)$ we get the following expression for VIX squared:

$$\left(\frac{VIX_t}{100}\right)^2 = -\frac{2e^{r(T-t)}}{T-t}\underbrace{E\left[\sum_{d=t+1}^{T}\lambda_1 V_d + \lambda_0 h_d\middle|\mathcal{F}_t\right]}_{(\Delta)}. \tag{A.7}$$

In order to compute the quantity (Δ) in (A.7) we use the mathematical induction method. $\forall\, l = t, \ldots, T$ we assume that:

$$E\left[\sum_{d=t+1}^{T}\lambda_1 V_d + \lambda_0 h_d\middle|\mathcal{F}_l\right] = C(l;T) + D(l;T)h_{l+1} + \sum_{d=t+1}^{l}\lambda_1 V_d + \lambda_0 h_d \tag{A.8}$$

with $C(T;T) = 0$ and $D(T;T) = 0$. First, we notice that all the quantities on the right side of (A.8) are known given the information at time l.

1. Since V_t and h_t are respectively adapted and predictable process our assumption is true for $l = T$ if $C(T;T) = 0$ and $D(T;T) = 0$.
2. We suppose the relation holds at time $l+1$ and we prove it for time l using the property of iterated expectations.

$$E\left[\sum_{d=t+1}^{T} \lambda_1 V_d + \lambda_0 h_d \middle| \mathcal{F}_l\right] = E\left[E\left[\sum_{d=t+1}^{T} \lambda_1 V_d + \lambda_0 h_d \middle| \mathcal{F}_{l+1}\right]\middle| \mathcal{F}_l\right]. \tag{A.9}$$

The quantity on the right hand side of Eq. (A.9) is equal to:

$$E\left[C(l+1;T) + D(l+1;T)h_{l+2} + \sum_{d=t+1}^{l+1} \lambda_1 V_d + \lambda_0 h_d \middle| \mathcal{F}_l\right]. \tag{A.10}$$

Substitute in (A.10) the definition of h_{l+2} and get:

$$C(l+1;T) + \alpha_0 D(l+1;T) + (\beta D(l+1;T) + \lambda_0)h_{t+1} + \textstyle\sum_{d=t+1}^{l}(\lambda_1 V_d + \lambda_0 h_d)$$
$$+E\left[(\alpha_1 D(l+1;T) + \lambda_1)\, V_{l+1} \middle| \mathcal{F}_l\right].$$

From (7.5) we get:

$$C(l+1;T) + \alpha_0 D(l+1;T) + \left[(\lambda_0 + \lambda_1 g(\theta)) + (\beta + \alpha_1 g(\theta))\, D(l+1;T)\right] h_{t+1}$$
$$+ \textstyle\sum_{d=t+1}^{l} \lambda_1 V_d + \lambda_0 h_d$$

and, by comparison with (A.8), we get the following system:

$$\begin{cases} C(l;T) = C(l+1;T) + D(l+1;T)\alpha_0 \\ D(l;T) = [\lambda_1 g(\theta) + \lambda_0] + (\alpha_1 g(\theta) + \beta)\, D(l+1;T) \end{cases} \tag{A.11}$$

with final conditions $C(T;T) = 0$ and $D(T;T) = 0$.

We show that if the following two conditions are satisfied:

- $\alpha_1 g(\theta) + \beta < 1$
- $\lambda_1 g(\theta) + \lambda_0 \leq 0$

the right hand of the Eq. (7.12) is positive, coherently with the fact of being equal to the squared VIX value. We notice that $D(l;T)$ is a linear difference equation whose solution at time $l = t$, $\forall t \leq T$ is given by:

$$D(t;T) = \underbrace{[\lambda_1 g(\theta) + \lambda_0]}_{\leq 0} \underbrace{\frac{1 - [\alpha_1 g(\theta) + \beta]^{T-t}}{1 - [\alpha_1 g(\theta) + \beta]}}_{>0}. \tag{A.12}$$

The solution (A.12) and the positivity of α_0 imply negative values for $C(t;T)$:

$$C(t;T) = \underbrace{C(T;T)}_{=0} + \underbrace{D(T;T)}_{=0} + \alpha_0 \sum_{l=t+1}^{T-1} \underbrace{D(l;T)}_{<0}$$
$$= \alpha_0 \left[\lambda_1 g(\theta) + \lambda_0\right] \left\{ \frac{T - t - 1 - [\alpha_1 g(\theta) + \beta] \frac{1-[\alpha_1 g(\theta)+\beta]^{(T-t)-1}}{1-[\alpha_1 g(\theta)+\beta]}}{1 - [\alpha_1 g(\theta) + \beta]} \right\}.$$

A.4 VIX Index: Autoregressive Model

In Eq. (7.6), we substitute the expression for h_{t+1} and h_t using the VIX adjusted as in (7.14). We obtain

$$\frac{VIX_t^{adj} - C_{30}}{D_{30}} = \alpha_0 + (\alpha_1 g(\theta) + \beta) \frac{VIX_{t-1}^{adj} - C_{30}}{D_{30}} + \alpha_1 (V_t - g(\theta) h_t) \Rightarrow$$
$$VIX_t^{adj} = \alpha_0 D_{30} + C_{30} \left[1 - (\alpha_1 g(\theta) + \beta)\right] + (\alpha_1 g(\theta) + \beta)\, VIX_{t-1}^{adj} + \alpha_1 D_{30} (V_t - g(\theta) h_t).$$

We can easily observe that VIX_t^{adj} is an $AR(1)$ and it can be written as:

$$VIX_t^{adj} = int + slope VIX_{t-1}^{adj} + \tau_t.$$

Trivially we have:

$$\begin{aligned} int &= \alpha_0 D_{30} + C_{30} \left[1 - (\alpha_1 g(\theta) + \beta)\right] \\ slope &= (\alpha_1 g(\theta) + \beta) \\ \tau_t &= \alpha_1 D_{30} (V_t - g(\theta) h_t). \end{aligned}$$

Using the explicit solution (7.13) for C_{30} and D_{30} and by rearranging, we get a simple expression for *int*:

$$\begin{aligned} int &= \alpha_0 \left(\lambda_1 g(\theta) + \lambda_0\right) \tfrac{1-slope^{30}}{1-slope} + \alpha_0 \left(\lambda_1 g(\theta) + \lambda_0\right) \left(29 - slope \tfrac{1-slope^{29}}{1-slope}\right) \\ &= 30 \alpha_0 \left(\lambda_1 g\left(\theta\right) + \lambda_0\right). \end{aligned}$$

References

O. Barndorff-Nielsen, Exponentially decreasing distributions for the logarithm of particle size. R. Soc. Lond. Proc. Ser. A **353**, 401–419 (1977)

O.E. Barndorff-Nielsen, Normal inverse gaussian distributions and stochastic volatility modelling. Scand. J. Stat. **24**(1), 1–13 (1997)

O.E. Barndorff-Nielsen, N. Shephard, Normal modified stable processes. Economics series working papers 072, Department of Economics, University of Oxford (2001)

O.E. Barndorff-Nielsen, J. Kent, M. Sørensen, Normal variance-mean mixtures and z-distributions. Int. Stat. Rev. **50**, 145–159 (1982)

F. Bellini, L. Mercuri, Option pricing in GARCH models. Technical report 124, Universitá di Milano Bicocca, Apr 2007

F. Bellini, L. Mercuri, Option pricing in a dynamic variance gamma model. J. Financ. Decis. Mak. **7**(1), 37–51 (2011)

F. Black, M.S. Scholes, The pricing of options and corporate liabilities. J. Polit. Econ. **81**(3), 637–54 (1973)

P. Carr, D.B. Madan, Option valuation using the fast fourier transform. J. Comput. Financ. **2**, 61–73 (1999)

Chicago Board Options Exchange, VIX-CBOE volatility index (2003). http://www.cboe.com/micro/vix/vixwhite.pdf

Chicago Board Options Exchange, CBOE skew index – skew (2011). http://www.cboe.com/micro/skew/introduction.aspx

P. Christoffersen, S. Heston, K. Jacobs, Option valuation with conditional skewness. J. Econ. **131**(1–2), 253–284 (2006)

P. Christoffersen, K. Jacobs, C. Ornthanalai, GARCH option valuation: theory and evidence. CREATES research papers 2012–2050, School of Economics and Management, University of Aarhus, May 2012

R. Cont, P. Tankov, *Financial Modelling with Jump Processes*. Financial Mathematics Series (Chapman & Hall/CRC, Boca Raton, 2003)

K. Demeterfi, E. Derman, M. Kamal, J. Zou, A guide to volatility and variance swaps. J. Deriv. **6**, 385–407 (1999)

J.C. Duan, The GARCH option pricing model. Math. Financ. **5**(1), 13–32 (1995)

J.C. Duan, J.G. Simonato, Empirical martingale simulation for asset prices. Manag. Sci. **44**(9), 1218–1233 (1998)

J.C. Duan, G. Gauthier, C. Sasseville, J.G. Simonato, An analytical approximation for the GARCH option pricing model. J. Comput. Financ. **2**(4), 75–116 (1999)

J.C. Duan, G. Gauthier, C. Sasseville, J.G. Simonato, Approximation the GJR-GARCH and the EGARCH option pricing models analytically. J. Comput. Financ. **9**(3), 41–69 (2006)

E. Eberlein, K. Prause, The generalized hyperbolic model: financial derivatives and risk measures, in *Mathematical Finance: Bachelier Congress 2000*, ed. by H. Geman, D. Madan, S.R. Pliska, T. Vorst. Springer Finance (Springer, Berlin/Heidelberg, 2002)

P. Embrechts, C. Klüppelberg, T. Mikosch, *Modelling Extremal Events for Insurance and Finance* (Springer, New York, 1997)

R. Engle (ed.), *ARCH: Selected Readings* (Oxford University Press, Oxford/New York, 1995). http://EconPapers.repec.org/RePEc:oxp:obooks:9780198774327

W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 3 (Wiley, New York, 1968)

K.R. French, Stock returns and the weekend effect. J. Financ. Econ. **8**(1), 55–69 (1980)

J. Hao, J.E. Zhang, GARCH option pricing models, the CBOE VIX, and variance risk premium. J. Financ. Econ. **11**(3), 556–580 (2013)

S.L. Heston, A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev. Financ. Stud. **6**, 327–343 (1993)

S.L. Heston, S. Nandi, A closed-form GARCH option valuation model. Rev. Financ. Stud. **13**(3), 585–625 (2000)

S.M. Iacus, *Option Pricing and Estimation of Financial Models with R* (Wiley-Blackwell, Chicester, 2011)

J. Kanniainen, B. Lin, H. Yang, Estimating and using GARCH models with VIX data for option valuation. J. Bank. Financ. **43**(C), 200–211 (2014)

Q. Liu, S. Guo, G. Qiao, VIX forecasting and variance risk premium: a new GARCH approach. N. Am. J. Econ. Financ. **34**, 314–322 (2015)

D.B. Madan, E. Seneta, The variance gamma (V.G.) model for share market returns. J. Bus. **63**(4), 511–524 (1990)

B. Mandelbrot, The variation of certain speculative prices. J. Bus. **36**, 394 (1963)

L. Mercuri, Option pricing in a GARCH model with tempered stable innovations. Financ. Res. Lett. **5**(3), 172–182 (2008)

R.C. Merton, Option pricing when underlying stock returns are discontinuous. J. Financ. Econ. **3**(1–2), 125–144 (1976)

C. Ornthanalai, A new class of asset pricing models with lévy processes: theory and applications. Technical report (2008)

W. Schoutens, *Lévy Processes in Finance* (Wiley, New York, 2003)

M.C.K. Tweedie, An index which distinguishes between some important exponential families, in *Statistics: Applications and New Directions*, ed. by J.K. Ghosh, J. Roy. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference (Indian Statistical Institute, Calcutta, 1984)

J.E. Zhang, Y. Zhu, VIX futures. J. Futur. Mark. **26**(6), 521–531 (2006)

Chapter 8
Optimal Adaptive Sequential Calibration of Option Models

Erik Lindström and Carl Åkerlindh

Abstract Option models needs to be recalibrated as new data becomes available. The updated model parameters will depend on previous parameters and new data, making adaptive sequential calibration a suitable choice. We introduce a method for optimally tuning the parameter adaptivity when non-linear filters are used for calibration, as well as extending the dynamics of the parameters.

The adaptivity is optimized by defining a statistical model, including both the option models and the adaptivity parameters. It turns out the corresponding (log-)likelihood function can be optimized through the EM algorithm, which ensures that the optimization is robust.

We evaluate the method on simulated data and S&P 500 index options, seeing that we can track variations in the model parameters well. The likelihood framework is also used for model selection where we find support for both complex option models as well as non-trivial adaptivity. This is made feasible with the optimal tuning presented in this chapter.

Keywords Unscented Kalman filter • EM algorithm • Sequential option calibration • Fourier Gauss-Laguerre option pricing

8.1 Introduction

The financial turnover in 2012 of equity linked derivatives was 6251 billion USD, while the turnover on FX derivatives markets was even greater with 67,358 billion USD, see Lindström et al. (2015). This means that even small errors in a model, or in the handling of that model, will have a serious economic impact.

There are nowadays an abundance of very capable option valuation models, see e.g. Cont and Tankov (2004) or Hirsa (2012) for overviews. Far less attention has been given to the statistical problem of calibrating these models to market data.

E. Lindström (✉) • C. Åkerlindh
Centre for Mathematical Sciences, Lund University, Lund, Sweden
e-mail: erikl@maths.lth.se; carl.akerlindh@matstat.lu.se

G. Consigli et al. (eds.), *Handbook of Recent Advances in Commodity and Financial Modeling*, International Series in Operations Research & Management Science 257,
DOI 10.1007/978-3-319-61320-8_8

Local volatility models, see Derman and Kani (1994) and Dupire (1994), were designed to provide a perfect in-sample fit to data. However, the out-of-sample performance is often less than stellar, cf. Dumas et al. (1998). One problem that may occur is that the recovered local volatility surface can induce arbitrage. Lindholm (2014a) provides an elegant, but computationally demanding solution, where the local volatility surface is forced to fulfill certain PDE constrains that eliminates the arbitrage opportunities. Dumas et al. (1998) showed that local volatility model rarely outperforms a modified Black-Scholes model, while Lindholm (2014b) showed that an ordinary Heston stochastic volatility model outperforms a tweaked Black-Scholes model when in terms of quadratic hedging. This view is also shared by Alexander and Kaeck (2012) who concludes that more advanced models does in fact provide better hedges than simpler models. We are therefore focusing on parametric models in this chapter.

The statistical problem of calibrating models to market data becomes increasingly complicated when more and more complex models are being used. A single parameter can be found using simple methods like line search, but this is not feasible in a more general context. The calibration problem may also be multi-modal for high dimensional data sets. The expectation maximization (EM) algorithm is one popular solution to these problems, cf. Rydén (2008).

The market practice when it comes to calibration is some kind of least squares minimization of the difference between market prices and model implied prices, typically minimizing the squared difference between todays market prices and corresponding model prices, see Hull (2009). Forbes et al. (2007), Kim and Singleton (2012), and Hurn et al. (2015) use filtering techniques to recover constant model parameters and risk premiums. Another approach was taken in Lindström et al. (2008) and Lindström and Jingyi (2013), where a non-linear filter was used to calibrate the model to market data using time varying model parameters. That approach generally led to more robust estimates and better out-of-sample forecasts. However, this comes at the cost of some tuning matrices, which in Lindström et al. (2008) were manually tuned.

The contribution of this chapter is twofold. The first is to derive a statically sound and numerically robust framework for estimating these tuning matrices. This is done by combining non-linear filtering and smoothing techniques with the EM algorithm introduced in Dempster et al. (1977) in a general context and by Shumway and Stoffer (1982) in a time series context. The EM algorithm provides closed form estimates for the tuning matrices, and is generally more robust than direct likelihood maximizing procedures for many models, see e.g. Rydén (2008) and Regland and Lindström (2012). The output of the algorithm will converge towards the maximum likelihood (ML) estimate of the tuning matrix, cf. Lindström (2013), ensuring optimality of our procedure. Our second contribution is that we extend the dynamics in Lindström et al. (2008) from one type to three different types of parameter dynamics. All of these share enough features so that the framework derived in this chapter is applicable.

We will throughout the chapter work with European call options, either directly or by converting European put options into call options through the put-call parity whenever the put options are more liquid. We evaluate the methodology on several sets of simulated data as well as on S&P 500 index options from 2004 to 2008.

The remainder of the chapter is organized as follows. Section 8.2 reviews some common calibration methods, including non-linear filters, and derives a framework for tuning the filters by using formal statistical tools. The tuning strategy will be evaluated in Sect. 8.3 on simulated data and in Sect. 8.4 on S&P 500 index option data. Section 8.5 concludes the chapter.

8.2 Methods

We start by stating some facts about data that are well known, yet often neglected. One of these is the lack of a single price. What is available are bid prices $\pi_t^{Bid}(K_i, \tau_i)$ and ask prices $\pi_t^{Ask}(K_i, \tau_i)$, as well as closing prices. A trivial consequence is that any predicted option value between the bid and the ask price does not introduce arbitrage.

The most common proxy for a single market price is the mid price which is defined as the average of the bid and ask prices

$$\pi_t^\star(K_i, \tau_i) = \frac{\pi_t^{Bid}(K_i, \tau_i) + \pi_t^{Ask}(K_i, \tau_i)}{2}. \tag{8.1}$$

Requiring that the model should provide a perfect fit to the mid price is most likely going to result in a heavy overfit of the data.

Another real world problem is that calibrated parameters tends to change over time (i.e. more than what is motivated by the randomness due to the bid-ask spread). This is in contrast to the intended role of the parameters in the underlying probabilistic models. Practitioners typically interpret this empirical fact that any serious calibration procedure must be adaptive (seeing the parameters as locally constant instead of globally), yet offer some predictive performance.

8.2.1 Review of Calibration Methods

The most popular calibration method today is weighted least squares (WLS), see Bates (1996), Cont and Tankov (2004), and Hull (2009), which is defined the parameter vector $\hat{\theta}$ that minimizes the weighted sum of squared differences between the observed mid price $\pi^\star$ and the corresponding model price π,

$$\hat{\theta}_t = \arg\min_{\theta\in\Theta} \sum_{s=t_0}^{t} \sum_{i=1}^{N_s} \lambda_{s,i} \left(\pi_s^\star(K_i, \tau_i) - \pi_s(K_i, \tau_i; \theta)\right)^2. \tag{8.2}$$

It is common to choose the weights $\lambda_{s,i}$ as a constant or proportional to the inverse of the squared bid-ask spread (the latter is optimal from a statistical point of view), thereby relating the size of the bid-ask spread to the quality of the quoted prices,

$$\lambda_{s,i} \propto \frac{1}{\left(\pi_s^{Bid}(K_i, \tau_i) - \pi_s^{Ask}(K_i, \tau_i)\right)^2}. \tag{8.3}$$

It is often argued that a perfect fit to market data is crucial, and this translates into using only the most recent data, i.e. $t_0 = t$. However, this choice brings some unpleasant philosophical implications. An implicit assumption is made that old data does not improve the estimation results, as the summation is restricted to current data. That means that the past is of limited use to predict the current prices, and hence that currently available information is of limited use for predictions of the future! It is therefore hardly surprising that the resulting estimates are noisy, see Lindström et al. (2006) and Lindström and Jingyi (2013). Some remedy is found by adding a (Tikhonov type) regularization term, see Cont and Tankov (2004), but this means that the perfect fit paradigm must be abandoned. Another alternative would be to consider exponential forgetting type algorithms, see e.g. Nystrup et al. (2016), but that often leads to ad hoc algorithms, rather that statistically sound methods.

An approach based on non-linear filtering was introduced in Lindström et al. (2006, 2008). This inspiration can be traced to stochastic volatility models, where the latent volatility state can be recovered using a non-linear filter, e.g. extended Kalman filter (EKF), unscented Kalman filter (UKF), ensemble Kalman filter (EnKF) or particle filter. The statistical model would then take a slightly different form

$$\pi_t^{\star}(K_i, \tau_i) = \pi_t(K_i, \tau_i; \theta, V_t) + \varepsilon_t, \tag{8.4a}$$

$$V_t \sim p(V_t | V_{t-1}; \theta), \tag{8.4b}$$

where ε_t is a N_t dimensional zero mean random vector with covariance matrix $\mathbf{V}[\varepsilon] = \mathbf{R}$. A possible design of $\mathbf{R}$ is to take it as a diagonal matrix, as the noise due to the bid-ask spread is assumed to be independent between observations. Hence, we could assign diagonal elements as $\mathbf{R}_{ii} = c \cdot \lambda_{t,i}$, with $\lambda_{t,i}$ being defined in Eq. (8.3). While $\lambda_{t,i}$ takes care of the scaling related to the size of the bid-ask spread, c is related to the distributional assumption on the measurement noise. Two common assumptions is that the real price is uniform inside the bid-ask spread, or that the bid ask spread is a confidence interval covering the real price with some probability. If the measurement error is assumed to be uniformly distributed c is chosen as $c = 1/12$ which is the variance of a standard uniform random variable. If the measurement error is assumed to be normally distributed and the bid-ask spread is assumed to be a 95% confidence interval (approximately 4 standard deviations wide) c is selected as $c = 1/4^2$. Another alternative is to estimate $\mathbf{R}$ from the data.

The model specified by Eq. (8.4) does not provide an easy method for estimating the parameters. However, this can be done by including the parameters in the latent state equation, see Åström and Eykhoff (1971) for some early results and Lindström et al. (2012) and Ionides et al. (2015) for recent results on convergence of this type of estimators. The sequential (but non-adaptive) calibration model is given as

$$\pi_t^\star(K_i, \tau_i) = \pi_t(K_i, \tau_i; \theta_t, V_t) + \varepsilon_t, \tag{8.5a}$$

$$\theta_t = \theta_{t-1}, \tag{8.5b}$$

$$V_t \sim p(V_t | V_{t-1}; \theta_{t-1}). \tag{8.5c}$$

This is a dynamic Bayesian method, but does not address the problem of time varying parameters. Adaptive estimation is achieved if the state space model is reformulated as a self-organizing state space model, adding random walk dynamics to the latent parameter vector, cf. Anderson and Moore (1979) and Kitagawa (1998). This defines the model as

$$\pi_t^\star(K_i, \tau_i) = \pi_t(K_i, \tau_i; \theta_t, V_t) + \varepsilon_t, \tag{8.6a}$$

$$\theta_t = \theta_{t-1} + \eta_t, \tag{8.6b}$$

$$V_t \sim p(V_t | V_{t-1}; \theta_{t-1}), \tag{8.6c}$$

where $\mathbf{V}[\eta] = \mathbf{Q}$ is the covariance matrix of the latent parameter vector error.

It can be argued that the latent volatility evolves according to the historical measure. However, the estimated risk premiums reported in Forbes et al. (2007) was not statistically significant, although later studies such as Hurn et al. (2015) did. The adaptive estimation methodology introduced in this chapter will in general result in wider confidence intervals for the parameter estimates. We therefore anticipated that it would be unlikely to obtain statistical significance of the risk premiums and hence assume it is zero. That implies that we approximate the historical measure with the risk neutral measure.

The model defined in Eq. (8.6) recasts the calibration into a non-linear filtering problem. There are many good algorithms available, both deterministic filters (EKF or UKF) or stochastic versions (EnKF or particle filters). The additional cost is the introduction of a tuning matrix $\mathbf{Q}$. It was chosen as a diagonal matrix in Lindström et al. (2008), with some manual tuning of the elements.

Our experience is that the non-linear filtering should not be applied directly to the standard formulation of the option valuation model, but rather to a model where the parameters has been transformed so that the random walk can not hit the boundary of the parameter space. For the parameters with positive constraint we use the natural logarithm, and for the correlation parameter we use the inverse hyperpolic tangent.

The results in Lindström et al. (2008) indicate that a reasonably well tuned filter performs well when calibrating option models to market data. However, it is known that some models are difficult to calibrate as there are many parameters, and possibly competing features in the models. A structured methodology for finding a near optimal $\mathbf{Q}$ matrix would therefore expand the class of statistically feasible models.

A minor modification was introduced in Lindström and Jingyi (2013) where the underlying asset was included as well. Doing so lead to a very minor deterioration of the performance of the filter, but the upside was that quadratic hedging strategies were obtained for free as the filter is computing all expected values, variances and covariances needed for the hedging strategy.

8.2.2 Extended Parameter Dynamics

The results in Lindström et al. (2008) indicates that the parameter are changing over time, some are varying around a fixed level while others are trending. We are therefore considering alternative adaptive dynamics.

8.2.2.1 Random Walk Dynamics

A simple method for introducing varying parameters was intruduced by Samuelson (1965) who argued that the parameter vector could be modeled as a random walk (RW). Hence, the parameters are given as

$$\theta_t = \theta_{t-1} + \varepsilon_t \tag{8.7}$$

where $\varepsilon \sim N(0, \mathbf{Q})$. This parameter dynamics was used in Lindström et al. (2008), cf. Equation (8.6), where it was shown that it could follow trends in the parameters. Using this dynamics however, introduces $d(d+1)/2$ tuning constants, where d is the dimension of the parameter vector, recalling that the covariance matrix is symmetric.

8.2.2.2 Random Coefficient Dynamics

Another simple dynamics is the random coefficient (RC) dynamics that was used in Schaefer et al. (1987). They treat the parameter at different time points as *iid* Gaussian random variables with fixed mean μ. The dynamics of the parameter vector is then given by

$$\theta_t = \mu + \varepsilon_t \tag{8.8}$$

where $\varepsilon \sim N(0, \mathbf{Q})$. The RC dynamics allows the parameters to vary (compared to fixed parameters), but not to follow trends. This dynamics introduces d additional tuning constants in comparison to RW, for a total of $d(d+1)/2 + d$.

8.2.2.3 Mean Reversion Dynamics

A third alternative that encompasses both the RW and the RC dynamics as special cases is the mean reversion (MR) dynamics discussed in Rosenberg (1973). The idea is to model the parameters as a stationary $VAR(1)$ process

$$\theta_t = \mu + A\theta_{t-1} + \varepsilon_t \tag{8.9}$$

where $\varepsilon \sim N(0, \mathbf{Q})$. The RC dynamics is recovered when $A = 0$, while the RW dynamics is recovered when $\mu = 0$ and $A = \mathbf{I}_d$ is a d-dimensional identity matrix. The matrix A introduces d^2 more constants to tune compared to the RC dynamics, resulting in a total of $d(d+1)/2 + d + d^2$ tuning constants.

8.2.3 *Optimal Tuning*

All of the proposed parameter dynamics introduce tuning constants that affect the behaviour of the filter. We will denote the set of tuning parameters with Φ, which in the case of MR is given by $\Phi = \{A, \mu, \mathbf{Q}\}$. In this section we will derive expressions for the optimal tuning of the parameter dynamics presented in Sect. 8.2.2.

We start with the model presented in Eq. (8.6), substituting the parameter process with the appropriate dynamics. The first important property to note is that the latent process V_t and the parameter process θ_t simultaneously satisfies the Markov property. Assuming independence of the latent process and the parameter process, we could see this as a hidden Markov model, and we then get the joint likelihood function

$$p_\Phi(V_{0:T}, \theta_{0:T}, \pi^\star_{1:T}) = p(V_0, \theta_0) \prod_{t=1}^{T} p(\pi^\star_t | V_t, \theta_t) p_\Phi(\theta_t | \theta_{t-1}) p(V_t | V_{t-1}, \theta_{t-1}), \tag{8.10}$$

where it is important to note that only $p_\Phi(\theta_t|\theta_{t-1})$ depends on the tuning parameters. Our goal is to find an estimate for the tuning parameters, given all observed prices $\pi^\star_{1:T}$.

By making an initial guess Φ_k of the tuning parameters, the EM algorithm iteratively maximizes the expectation of the joint log-likelihood given the observations. It can be shown that the EM algorithm increases the value of the likelihood function with each iteration, see Dempster et al. (1977). It is generally considered to be a robust and typically a rapidly converging method, cf. Rydén (2008). In other words, the estimates of the tuning parameters are gradually improved by iteratively maximizing the objective function given by

$$Q(\Phi|\Phi_k) = \mathbf{E}\left[\log p_\Phi(V_{0:T}, \theta_{0:T}, \pi^\star_{1:T}) | \pi^\star_{1:T}, \Phi_k\right]. \tag{8.11}$$

The construction of this function is usually refered to as the E-step. Computing the optimal updated set of tuning parameters is then done in the M-step, given by

$$\begin{aligned}\Phi_{k+1} &= \arg\max_{\Phi} Q(\Phi|\Phi_k) \\ &= \arg\max_{\Phi} \mathbf{E}\left[\log p_{\Phi}(\theta_t|\theta_{t-1})|\pi^{\star}_{1:T}, \Phi_k\right].\end{aligned} \tag{8.12}$$

Since θ_t is a Gaussian process for all of the proposed dynamics, the maximizing values are given by solving quadratic equations, which can be done analytically. This makes the EM algorithm a very good method of acquiring near optimal estimates, which converges to the ML estimates as the number of iterations tend to infinity.

However, when new data becomes available, we rarely need to iterate more than one or two times when initializing the algorithm from an already tuned parameter set Φ. We believe that the set of tuning parameters Φ is much more persistent than the actual parameters θ. This means that we do not have to update Φ every time new data becomes available.

In order to compute the EM estimates we need to obtain filter estimates and variances defined as

$$\theta_{t|t} = \mathbf{E}\left[\theta_t|\pi^{\star}_{1:t}\right], \tag{8.13a}$$

$$P_{t|t} = \mathbf{V}\left[\theta_t|\pi^{\star}_{1:t}\right], \tag{8.13b}$$

and smoothing estimates and variances defined as

$$\theta_{t|T} = \mathbf{E}\left[\theta_t|\pi^{\star}_{1:T}\right], \tag{8.14a}$$

$$P_{t|T} = \mathbf{V}\left[\theta_t|\pi^{\star}_{1:T}\right]. \tag{8.14b}$$

To compute these we use the non-linear Unscented Kalman filter and smoother, see Murata and Kashino (2013) for an overview, but other types of filters could also be applied.

For tuning the parameters introduced by RW, RC and MR dynamics, we state the following lemmas for optimal choices of EM updates, cf. Shumway and Stoffer (1982).

Lemma 2.1 (Random walk dynamics) *Given the random walk parameter dynamics in Eq.* (8.7)*, the optimal EM update of the tuning parameter is given by*

$$\begin{aligned}\mathbf{Q}_{k+1} = \frac{1}{T}\sum_{t=1}^{T}&(\theta_{t|T} - \theta_{t-1|T})(\theta_{t|T} - \theta_{t-1|T})^{\mathsf{T}} \\ &+ P_{t|T} + P_{t-1|T} - P_{t|T}L_{t-1}^{\mathsf{T}} - L_{t-1}P_{t|T},\end{aligned} \tag{8.15a}$$

where

$$L_t = P_{t|t}(P_{t|t} + \mathbf{Q}_k)^{-1} \tag{8.16}$$

is the Kalman smoother gain.

Lemma 2.2 (Random coefficient dynamics) *Given the random coefficient parameter dynamics in Eq. (8.8), the optimal EM updates of the tuning parameters are given by*

$$\mu_{k+1} = \frac{1}{T}\sum_{t=1}^{T}\theta_{t|T}, \tag{8.17a}$$

and

$$\mathbf{Q}_{k+1} = \frac{1}{T}\sum_{t=1}^{T}(\theta_{t|T} - \mu_{k+1})(\theta_{t|T} - \mu_{k+1})^{\intercal} + P_{t|T}. \tag{8.17b}$$

Lemma 2.3 (Mean reversion dynamics) *Given the mean reverting parameter dynamics in Eq. (8.9), the optimal EM updates of the tuning parameters are given by*

$$\begin{aligned}\left[\mu_{k+1}\ A_{k+1}\right] = &\left[\textstyle\sum_{t=1}^{T}\theta_{t|T}\ \sum_{t=1}^{T}\theta_{t|T}\theta_{t-1|T}^{\intercal} + P_{t|T}L_{t-1}^{\intercal}\right]\\ &\times\begin{bmatrix}\sum_{t=1}^{T}1 & \sum_{t=1}^{T}\theta_{t-1|T}^{\intercal}\\ \sum_{t=1}^{T}\theta_{t-1|T} & \sum_{t=1}^{T}\theta_{t-1|T}\theta_{t-1|T}^{\intercal} + P_{t-1|T}\end{bmatrix}^{-1},\end{aligned} \tag{8.18a}$$

and

$$\begin{aligned}\mathbf{Q}_{k+1} = \frac{1}{T}\sum_{t=1}^{T}&(\theta_{t|T} - A_{k+1}\theta_{t-1|T} - \mu_{k+1})(\theta_{t|T} - A_{k+1}\theta_{t-1|T} - \mu_{k+1})^{\intercal}\\ &+ P_{t|T} + A_{k+1}P_{t-1|T}A_{k+1}^{\intercal}\\ &- P_{t|T}L_{t-1}^{\intercal}A_{k+1}^{\intercal} - A_{k+1}L_{t-1}P_{t|T},\end{aligned} \tag{8.18b}$$

where

$$L_t = P_{t|t}A_k^{\intercal}(A_kP_{t|t}A_k^{\intercal} + \mathbf{Q}_k)^{-1} \tag{8.19}$$

is the Kalman smoother gain.

The measurement noise variance is not always known, in which case there are two options, guessing or estimating. Assuming a structure of the measurement noise variance as $\mathbf{R} = R \cdot \mathbf{I}_{N_t}$ we can estimate the scaling factor R at the same time as we estimate all other tuning constants. The expression for the optimal EM update is presented in the following lemma.

Lemma 2.4 (Measurement noise) *Assuming the noise of the measurements is* iid *Gaussian, i.e.* $\mathbf{R} = R \cdot \mathbf{I}_{N_t}$ *the optimal EM update of the scaling factor* R *is given by*

$$R_{k+1} = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N_t}\sum_{i=1}^{N_t}\left(\pi_t^{\star}(K_i,\tau_i) - \mathbf{E}[\pi_t(K_i,\tau_i;\theta_t,V_t)|\pi_{1:T}^{\star},R_k]\right)^2 + \mathbf{V}[\pi_t(K_i,\tau_i;\theta_t,V_t)|\pi_{1:T}^{\star},R_k]. \tag{8.20}$$

Estimating a measurement noise variance with this structure adds a single additional tuning constant.

8.2.4 Model Selection

An additional benefit of using the EM algorithm in combination with filtering and smoothing techniques is that we can compute the likelihood of the measurements. This makes it possible to compare the parameter dynamics with model selection criteria such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC). The dynamics with the lowest AIC value is the optimal choice for prediction, while the BIC is a consistent model selection criteria. We will use both of these to draw conclusions in our simulation and empirical study.

8.3 Simulations

To validate the performance of the tuning and model selection we simulate data from all parameter dynamics proposed in Sect. 8.2.2. We simulate 250 weeks, each consisting of totally 15 European call option prices, with maturity in three, 6 and 12 months and five strikes equidistantly spaced between 0.9 and 1.1 times the current spot price. To mimic the uncertainty related to the bid-ask spread, measurement errors are added to the simulated option prices. The errors are assumed to be normally distributed with a known variance $\mathbf{R} = 1/16 \cdot \mathbf{I}_{N_t}$. The standard Black-Scholes model, defined under the risk neutral measure $\mathbb{Q}$, is given by

$$\mathrm{d}S_t = rS_t\mathrm{d}t + \sigma S_t \mathrm{d}W_t \tag{8.21}$$

is used both for simulation and estimation, in combination with RW, RC and MR parameter dynamics. In Table 8.1 the dynamic specific parameters are shown, that are used for the simulations. Note that these describe the evolution of the transformed model parameters, i.e.

$$\theta_t = \begin{pmatrix}\log r_t \\ \log \sigma_t\end{pmatrix}. \tag{8.22}$$

Table 8.1 Parameter dynamics used for simulating from the Black-Scholes model. These are chosen to mimic those that are estimated for market data

Simulated	A	μ	$\mathbf{Q}$
RW	–	–	$\begin{pmatrix} 0.2 & 0 \\ 0 & 0.2 \end{pmatrix}$
RC	–	$\begin{pmatrix} -4.6 \\ -1.2 \end{pmatrix}$	$\begin{pmatrix} 0.2 & 0 \\ 0 & 0.2 \end{pmatrix}$
MR	$\begin{pmatrix} 0.9 & 0.01 \\ 0.03 & 0.9 \end{pmatrix}$	$\begin{pmatrix} -0.07 \\ -0.06 \end{pmatrix}$	$\begin{pmatrix} 0.03 & 0 \\ 0 & 0.5 \end{pmatrix}$

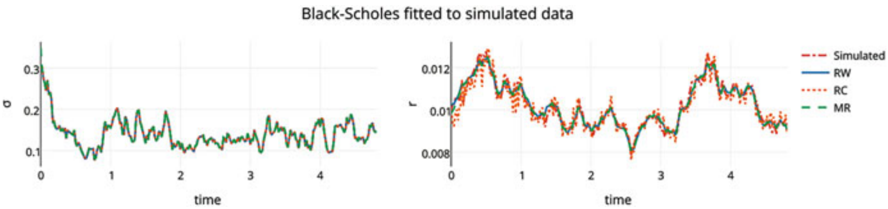


Fig. 8.1 Estimated and real parameters (volatility σ *left*, interest rate r *right*) when calibrating to simulated Black-Scholes data with mean reverting parameters

The model and tuning parameters is estimated using all three parameter dynamics on each data set. We iterate the EM algorithm 100 times, which is more than enough to reach convergence. We find that our method delivers far better results (in terms of the log-likelihood) than what a manually tuned model would obtain, see Lindström and Åkerlindh (2014). Finally, we discard the first 20 weeks as a burn-in when computing the filter and EM estimates, in order to reduce the influence of the initial values.

In Fig. 8.1 we see the parameter paths with RW, RC and MR dynamics estimated on data generated with mean reverting parameters. We see that all dynamics track the volatility very well, while the less informative (and thus less influential on prices) rate parameter is harder to track perfect. It is also clear that the RC dynamics result in noisier estimates than RW and MR.

In order to determine which model fits best in each case, we calculate the AIC and BIC. We recall that BIC is the consistent model selection criteria while AIC selects the model that provide the best forecasts. The results of the simulation study are presented in Table 8.2. In all three cases the correct model is selected by the BIC, while AIC always selects MR dynamics, regardless of the actual dynamics. The results of the simulation study validates the effectiveness of not only the tuning, but also the model selection properties of our method. The MR dynamics is usually a good choice regardless of the actual dynamics, as the drawbacks of the overfitting and additional computational cost is more than made up for by the forecasting properties.

Table 8.2 Comparison of models estimated using the RW, RC and MR dynamics when applied to simulated data from each model. The consistent model selection criteria BIC selects the correct model in all three cases, while the AIC selects MR in all cases as the dynamics that provide the best forecasts

Simulated	Estimated	ℓ	AIC	BIC
RW	RW	−1148.94	2303.88	**2314.22**
	RC	−2264.94	4539.89	4557.12
	MR	−1139.98	**2297.96**	2328.98
RC	RW	−1460.03	2926.06	2936.40
	RC	−1305.77	2621.55	**2638.78**
	MR	−1296.24	**2610.47**	2641.49
MR	RW	−1318.50	2643.01	2653.35
	RC	−1850.42	3710.84	3728.07
	MR	−1281.71	**2581.42**	**2612.44**

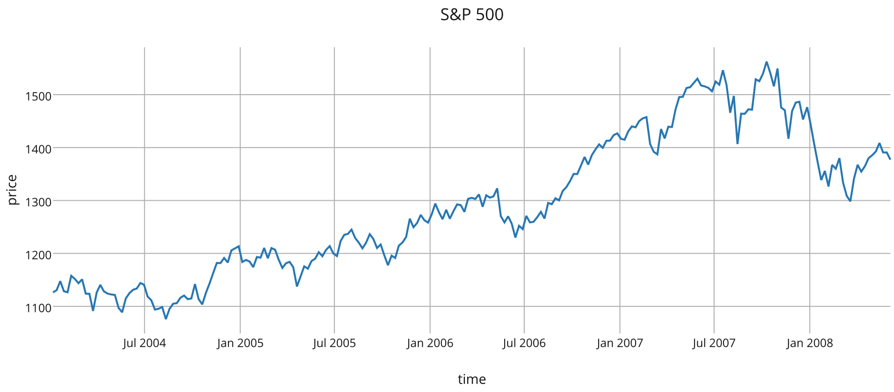


Fig. 8.2 S&P 500 index from early 2004 to mid 2008

8.4 Empirical Study

We also evaluate our method on S&P 500 index options from the beginning of 2004 to the middle of 2008. This data set was used in Poulsen et al. (2009). The first part of the data set is calm, while the second part includes the opening stages of the financial crisis. The index is presented in Fig. 8.2.

The option data set is identical to the one in our simulation studies, with options having three, 6 or 12 months to maturity and moneyness ranging equidistantly from 0.9 to 1.1.

We calibrate two models to the market data, the Black-Scholes model and the Heston stochastic volatility model, see Heston (1993). The Heston prices are computed using Fourier methods, more precisely a Fourier-Gauss Laguerre method with an adaptive choice of the integration path. A large number of methods, including Fourier, cosine, Monte Carlo, quasi-Monte Carlo and PDE methods were compared in von Sydow et al. (2015) where it was shown that this method is very

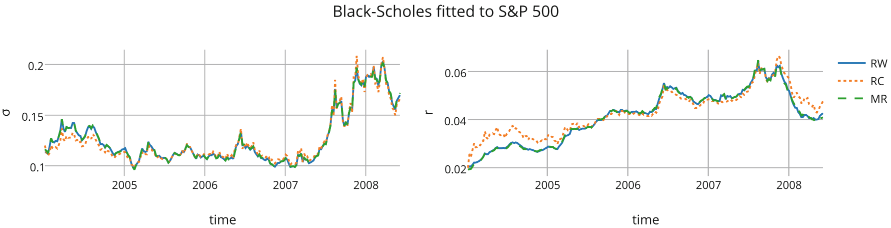


Fig. 8.3 Estimated parameters (volatility σ *left*, interest rate r *right*) when calibrating the Black-Scholes model to S&P 500 index options between 2004 and 2008

fast and outperforms all other methods in terms of accuracy. We believe that the errors when computing the prices must be dominated by the tolerance levels in the optimization, something we obtain with this method.

All calibrations in this section uses an estimate of the measurement error covariance matrix $\mathbf{R}$, of the form presented in Lemma 2.4. This gives much better results for the Black-Scholes model (where we can expect that the model misfit to be substantial), but no large improvement for the Heston model.

8.4.1 The Black-Scholes Model

The risk neutral dynamics for the well known Black-Scholes model is given in Eq. (8.21). We calibrate two parameters in this study, the volatility σ and the interest rate r. The resulting calibrated parameters, when using the RC (dotted line), RW (solid line) and MR (dashed line) are presented in Fig. 8.3. A visual inspection shows that the RW and MR dynamics results in similar parameters while the RC seems noisier than the other two.

The calibrated parameters reveals that the calm period between 2004 and 2007 indeed corresponded to a low volatility (around 10%) with steadily increasing interest rates as the bubble started to build up. The crisis changes the situation with noticeably higher volatilities and eventually much lower rates, as government interventions begins.

8.4.2 The Heston Model

We also calibrate the Heston model to the same set of market data. The risk neutral formulation of the Heston model is given by

$$dS_t = rS_t dt + \sqrt{V_t} S_t dW_t^{(1)}, \tag{8.23a}$$

$$dV_t = \kappa(\theta - V_t)dt + \sigma\sqrt{V_t} dW_t^{(2)}, \tag{8.23b}$$

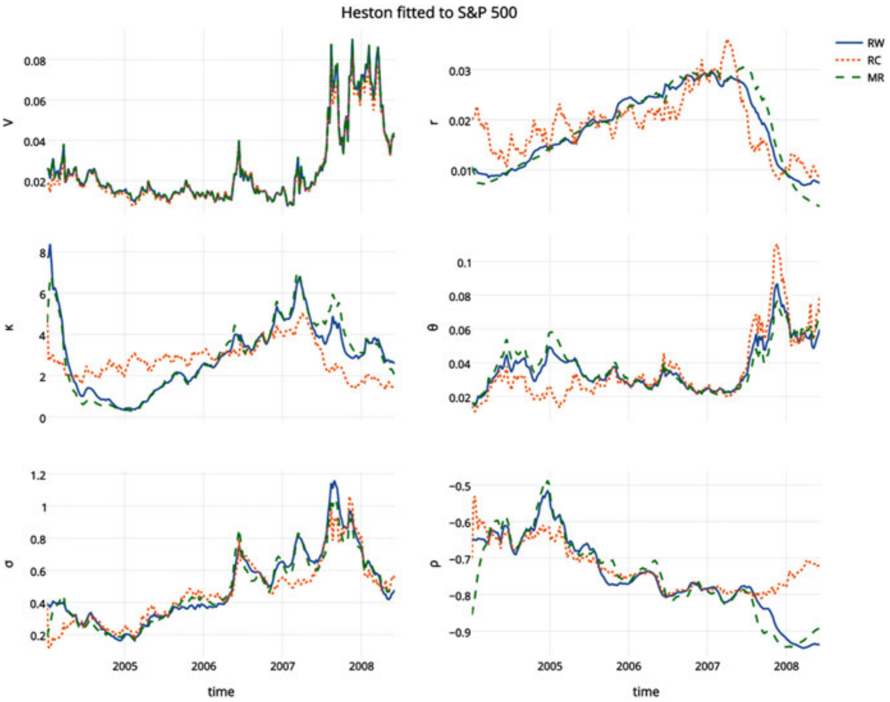


Fig. 8.4 Estimated parameters when calibrating the Heston model to S&P 500 index options using the RW (*solid line*), RC (*dotted line*) and MR (*dashed line*) parameter dynamics. The instantaneous volatility V is presented *top left*, the recovered interest rate r *top right*, mean reversion of the volatility κ *mid left*, the mean reversion level θ *mid right*, the vol-vol parameter σ *bottom left* and the correlation ρ, between the Brownian motions, *bottom right*

$$dW_t^{(1)} dW_t^{(2)} = \rho dt. \tag{8.23c}$$

where S is the index level, V is the latent stochastic volatility and ρ is the instantaneous correlation between the index and volatility, capturing the leverage effect often found in financial data, see Cont (2001) and Nystrup et al. (2015) for overviews over stylized facts. The latent volatility process is approximated using Euler discretization, and is filtered together with the model parameters.

The calibrate parameters for all three dynamics are presented in Fig. 8.4, where we again see that the MR and RW are reasonably similar while the RC dynamics results in different estimates.

It can be seen that the Heston model captures the varying volatility similarly to the Black-Scholes model, but we also find new information in Fig. 8.4 such that the correlation nearly breaks down (close to -1) during the crisis. That meant that delta-vega neutral risk management strategies did not work nearly as well as they used to do before the crisis.

Table 8.3 Comparison of the Black-Scholes and Heston models when estimated with RW, RC and MR parameter dynamics. The MR dynamics is selected by the AIC as the best forecasting dynamics for both models. We observe a great benefit of using MR parameter dynamics in the Heston model, while it is enough to use RW dynamics for the simpler Black-Scholes model

Model	Estimated	ℓ	AIC	BIC
BS	RW	−8293.94	16595.87	**16609.30**
	RC	−9076.02	18164.04	18184.18
	MR	−8281.60	**16583.20**	16616.76
Heston	RW	−3630.97	7293.94	7347.65
	RC	−5952.12	11946.24	12016.72
	MR	−3461.99	**7015.99**	**7170.39**

8.4.2.1 Numerical Results

One of the most appealing features of the (log-)likelihood based framework is that standard statistical tests can be carried out. We compute the AIC and BIC for all models and parameters dynamics for the data set. The results are presented in Table 8.3.

These results pretty much confirms the impressions from Figs. 8.3 and 8.4 with the RW and MR dynamics behaving similarly, while the RC dynamics lags behind. We see that while the RW is sufficient according to the BIC for the Black-Scholes model, it is not sufficient for the more complex Heston model. The MR parameter dynamics is better at describing the data when the complexity of the option valuation model increases. We anticipate that there will be more dependence between the parameters in a more complex model, and this is better captured by the MR dynamics, cf. Figs. 8.3 and 8.4.

Selecting the parameter dynamics based on the AIC favors the MR dynamics regardless of the model. This indicates that the MR parameter dynamics may be preferable for a large class of models if hedging is our primary purpose of the model.

The AIC does not require models to be nested, and can therefore be used to compared not only parameter dynamics, but also to compare the Black-Scholes model and the Heston model. Our analysis shows a very strong support for the Heston model in favor of the simpler Black-Scholes model.

8.5 Conclusion

Several studies have shown that the WLS technique is practically infeasible when the model complexity grows, while non-linear filters or penalized WLS work much better. This study introduced optimal tuning strategies for the sequential calibration technique introduced in Lindström et al. (2008), using three different parameter dynamics to better capture the structure of the underlying processes.

The optimal tuning introduced in this chapter resulted in better parameter estimates than a manually tuned filter. The framework also allowed us to extend the class of parameter dynamics. In other words, the introduction of optimal tuning strategies makes it possible to use more advanced models and parameter dynamics in practice. This is something we are convinced would lead to better hedging and risk management.

Our simulation study showed that the method which is based on the EM algorithm, is a fast and numerically stable way to estimate optimal tuning parameters. Convergence is typically reached in just a few iterations. It is possible to evaluate the likelihood for our extended model, and this was used for model selection, where we found that the correct underlying parameter dynamics was identified using the BIC.

It is likely that the optimal tuning parameters are persistent compared to the parameters. It would therefore make sense to update them less frequently than the actual parameters. For example, when used in practice the optimal tuning parameters could be reestimated weekly or monthly, while the sequential calibration of the option model is done daily.

References

C. Alexander, A. Kaeck, Does model fit matter for hedging? Evidence from FTSE 100 options. J. Futur. Mark. **32**(7), 609–638 (2012)

B.D.O. Anderson, J.B. Moore, *Optimal Filtering* (Prentice-Hall, Englewood Cliffs, 1979)

K.J. Åström, P. Eykhoff, System identification – a survey. Automatica **7**(2), 123–162 (1971)

D.S. Bates, Jumps and stochastic volatility: exchange rate processes implicit in Deutsche Mark options. Rev. Financ. Stud. **9**(1), 69–107 (1996)

R. Cont, Empirical properties of asset returns: stylized facts and statistical issues. Quant. Finan. **1**, 223–236 (2001)

R. Cont, P. Tankov, *Financial Modelling with Jump Processes* (CRC press, Boca Raton, 2004)

A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B Methodol. **39**(1), 1–38 (1977)

E. Derman, I. Kani, Riding on a smile. Risk **7**(2), 32–39 (1994)

B. Dumas, J. Fleming, R.E. Whaley, Implied volatility functions: empirical tests. J. Financ. **53**(6), 2059–2106 (1998)

B. Dupire, Pricing with a smile. Risk **7**(1), 18–20 (1994)

C.S. Forbes, G.M. Martin, J. Wright, Inference for a class of stochastic volatility models using option and spot prices: application of a bivariate kalman filter. Econ. Rev. **26**(2–4), 387–418 (2007)

S.L. Heston, A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev. Financ. Stud. **6**(2), 327–343 (1993)

A. Hirsa, *Computational Methods in Finance* (CRC Press, Hoboken, 2012)

J. Hull, *Options, Futures and Other Derivatives* (Pearson education, Upper Saddle River, 2009)

A.S. Hurn, K.A. Lindsay, A.J. McClelland, Estimating the parameters of stochastic volatility models using option price data. J. Bus. Econ. Stat. **33**(4), 579–594 (2015)

E.L. Ionides, D. Nguyen, Y. Atchadé, S. Stoev, A.A. King, Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. Proc. Natl. Acad. Sci. **112**(3), 719–724 (2015)

D.H. Kim, K.J. Singleton, Term structure models and the zero bound: an empirical investigation of Japanese yields. J. Econ. **170**(1), 32–49 (2012)

G. Kitagawa, A self-organizing state-space model. J. Am. Stat. Assoc. **93**(443), 1203–1215 (1998)
L. Lindholm, Calibration of local volatility surfaces under PDE constrains, in *Calibration and Hedging in Finance*. Phd Lic thesis, KTH Royal Institute of Technology (2014a)
L. Lindholm, Empirical performance of quadratic hedging strategies applied to European call options on an equity index, in *Calibration and Hedging in Finance*. Phd Lic thesis, KTH Royal Institute of Technology (2014b)
E. Lindström, Tuned iterated filtering. Stat. Probab. Lett. **83**(9), 2077–2080 (2013)
E. Lindström, C. Åkerlindh, Tuned sequential calibration of options, in *Euro Working Group for Commodities and Financial Modelling (EWGCFM)* (2014)
E. Lindström, E. Ionides, J. Frydendall, H. Madsen, Efficient iterated filtering. IFAC Proc. Vol. **45**(16), 1785–1790 (2012). ISSN:1474-6670, http://dx.doi.org/10.3182/20120711-3-BE-2027.00300
E. Lindström, G. Jingyi, Simultaneous calibration and quadratic hedging of options. Quantit. Qual. Anal. Soc. Sci. **7**(1) (2013)
E. Lindström, H. Madsen, J.N. Nielsen, *Statistics for Finance* (CRC Press, Boca Raton, 2015)
E. Lindström, J. Ströjby, M. Brodén, M. Wiktorsson, J. Holst, Calibration of option valuation models using sequential Monte Carlo methods, in *13th International Conference on Forecasting Financial Markets* (2006)
E. Lindström, J. Ströjby, M. Brodén, M. Wiktorsson, J. Holst, Sequential calibration of options. Comput. Stat. Data Anal. **52**(6), 2877–2891 (2008)
M. Murata, K. Kashino, Normalized unscented Kalman filter and normalized unscented RTS smoother for nonlinear state-space model identification, in *American Control Conference (ACC)* (IEEE, 2013), pp. 5462–5467
P. Nystrup, H. Madsen, E. Lindström, Stylised facts of financial time series and hidden Markov models in continuous time. Quant. Finan. (2015). doi:10.1080/14697688.2015.1004801
P. Nystrup, H. Madsen, E. Lindström, Long memory of financial time series and Hidden Markov models with time-varying parameters. J. Forecast. (2016). http://dx.doi.org/10.1002/for.2447
R. Poulsen, K.R. Schenk-Hoppé, C.-O. Ewald, Risk minimization in stochastic volatility models: model risk and empirical performance. Quant. Finan. **9**(6), 693–704 (2009)
F. Regland, E. Lindström, Independent spike models: estimation and validation. Czech J. Econ. Finan. (Finan. a uver) **62**(2), 180–196 (2012)
B. Rosenberg, Random coefficients models: the analysis of a cross section of time series by stochastically convergent parameter regression. Ann. Econ. Soc. Meas. **2**(4), 397–428 (1973)
T. Rydén, EM versus Markov chain Monte Carlo for estimation of hidden Markov models: a computational perspective. Bayesian Anal. **3**(4), 659–688 (2008)
P.A. Samuelson, Proof that properly anticipated prices fluctuate randomly. Ind. Manag. Rev. **6**(2), 41–49 (1965)
S. Schaefer, R. Brealey, S. Hodges, H. Thomas, *Alternative Models of Systematic Risk* (Graduate School of Business Studies, London, 1987)
R.H. Shumway, D.S. Stoffer, An approach to time series smoothing and forecasting using the EM algorithm. J. Time Ser. Anal. **3**(4), 253–264 (1982)
L. von Sydow, L.J. Höök, E. Larsson, E. Lindström, S. Milovanovic, J. Persson, V. Shcherbakov, Y. Shpolyanskiy, S. Siren, J. Toivanen, J. Waldén, M. Wiktorsson, M.B. Giles, J. Levesley, J. Li, C.W. Oosterlee, M.J. Ruijter, A. Toropov, Y. Zhao, BENCHOP: the benchmarking project in option pricing. Int. J. Comput. Math. **92**, 2361–2379 (2015). http://dx.doi.org/10.1080/00207160.2015.1072172

Chapter 9
Accurate Pricing of Swaptions via Lower Bound

Anna Maria Gambaro, Ruggero Caldana, and Gianluca Fusai

Abstract We propose a new lower bound for pricing European-style swaptions for a wide class of interest rate models. This method is applicable whenever the joint characteristic function of the state variables is either known in closed form or can be obtained numerically via some efficient procedure. Our lower bound involves the computation of a one dimensional Fourier transform independently of the underlying swap length. Finally the bound can be used as a control variate to reduce the confidence interval in the Monte Carlo simulation. We test our bound on different affine models, also allowing for jumps. The lower bound is found to be accurate and computationally efficient.

Keywords Pricing • Swaptions • Characteristic function • Fourier transform • Lower bound

9.1 Introduction

Libor based derivatives (swaps, caps, swaptions) are the most liquid derivatives traded in financial markets. In particular a European swaption is a contract that gives the right to its owner to enter into an underlying interest rate swap, i.e. it is an European option on a swap rate. It can be equivalently interpreted as an option on a portfolio of zero coupon bonds.

A.M. Gambaro (✉)
Dipartimento di Statistica e Metodi Quantitativi, Università Milano Bicocca, Milano, Italy
e-mail: a.gambaro@campus.unimib.it

R. Caldana
Accenture, Milano, Italy
e-mail: ruggero.caldana@gmail.con

G. Fusai
Faculty of Finance, Cass Business School, City University London, London, UK

Dipartimento di Studi per l'Economia e l'Impresa, Università del Piemonte Orientale "A. Avogadro", Novara, Italy
e-mail: Gianluca.Fusai.1@city.ac.uk

G. Consigli et al. (eds.), *Handbook of Recent Advances in Commodity and Financial Modeling*, International Series in Operations Research & Management Science 257,
DOI 10.1007/978-3-319-61320-8_9

Let t be the current date, T be the option expiration date, $T_1, \ldots, T_n$ be the underlying swap payment dates (by construction $t < T < T_1 < \ldots < T_n$) and R the fixed rate of the swap. The payoff of a receiver swaption is

$$payoff = \left(\sum_{h=1}^{n} w_h P(T, T_h) - 1 \right)^+ ,$$

where $w_h = R \cdot (T_h - T_{h-1})$ for $h = 1, \ldots, n-1$ and $w_n = R \cdot (T_n - T_{n-1}) + 1$, $P(T, T_h)$ is the price at time T of a zero coupon bond expiring at time T_h. The no-arbitrage fair price at time t is the discounted risk neutral expected value of the payoff

$$C(t, T, \{T_h\}_{h=1}^n, R) = \mathbb{E}_t \left[e^{-\int_t^T r(\mathbf{X}(s))ds} \left(\sum_{h=1}^{n} w_h P(T, T_h) - 1 \right)^+ \right], \qquad (9.1)$$

where $r(\mathbf{X}(s))$ is the short rate at time s, and $\mathbf{X}(s)$ denotes the state vector at time s of a multi-factor stochastic model.

Due to their importance and popularity, swaption market quotations are often used for calibration of interest rate models. However calibration procedure involves the pricing of a large number of swaptions (different option maturity, swap tenors and strikes), so the availability of an efficient pricing algorithm is required. Moreover Basel III accords introduced the Credit Value Adjustment (CVA) charge for over the counter contracts.[1] It is interesting to note that for the most simple and popular kind of interest rate derivative, i.e. interest rate swap, the (unilateral) C.V.A. can be estimated as a portfolio of forward start European swaptions.[2]

Since a closed-form formula of swaption price does not exist for many popular interest rate models, then several approximated pricing method have been developed in literature specially for affine interest rate models. The most important are those of Munk (1999), Collin-Dufresne and Goldstein (2002), Singleton and Umantsev (2002), and Schrager and Pelsser (2006).

[1]CVA of a contract price C(t) is the risk neutral expectation of the loss

$$CVA(t) := LDG \int_0^T \mathbb{E}[e^{-\int_0^t r(s)ds} max(C(t), 0)] \, d\mathbb{P}_D(t),$$

where LGD is the Loss Given Default and $\mathbb{P}_D(t)$ is the default probability in the interval $(t, t + dt)$. Risk adjusted price is: $C(t) - CVA(t)$.

[2]C.V.A. of an interest rate swap with payment dates $T_1, \ldots, T_n$ and fixed rate R, can be approximated by the following portfolio of swaptions:

$$CVA(t) \simeq LGD \sum_{i=1}^{n} \mathbb{P}_D(T_{i-1}, T_i) \cdot SWO(t, \{T_h\}_{i-1}^n, R).$$

.

Munk approximates the price of an option on a coupon bond by a multiple of the price of an option on a zero-coupon bond with time to maturity equal to the stochastic duration of the coupon bond. This method is fast but not very accurate for out of the money options.

The method of Collin-Dufresne and Goldstein is based on a Edgeworth expansion of the density of the swap rate and requires the calculation of the moments of the coupon bond. This procedure can be very time consuming.

Singleton and Umantsev (2002) introduce the idea of approximating the exercise region in the space of the state variables. This method has the advantage of producing precise results for a wide range of strikes (in particular for out of the money swaptions); however, it does not admit a simple extension to general affine interest rate models, because it requires the knowledge in closed form of the joint probability density function of the state variables. Moreover the Singleton and Umantsev method requires the calculus of as many Fourier inversion as the number of payment dates of the underlying swap contract. Hence the run time algorithm increases with the swap length.

Similarly to Singleton and Umantsev (2002), we propose a lower bound which is based on an approximation of the exercise region via an event set defined through a function of the model factors

$$\mathcal{G} = \{\omega \in \Omega : g(\mathbf{X}(\omega), T) \geq k\}, \tag{9.2}$$

where $\mathbf{X}$ is the vector of the model factors, Ω is the state space and g is a suitably chosen function approximating the exercise boundary.

Our pricing formula consists in the valuation of the option on the approximate exercise region and requires a single Fourier transform, performed with respect to the parameter k in formula (9.2).

The approximation we propose has several advantages. First of all it is a lower bound, so the direction of the error is known a priori, further it is very general as it can be applied to a wide class of models, provided that the characteristic function of the state variates is known (explicitly or numerically). It involves the computation of only one Fourier inversion, independently of the number of cash flows of the underlying swap. Finally, it can be used as a control variate to improve the accuracy of the Monte Carlo simulation method.

The paper is organized as follows. Section 9.2 introduces a general formula for lower bound on swaption prices, based on an approximation of the exercise region. Then we apply the general lower bound formula to the case of affine interest rate models and we find an efficient algorithm to calculate analytically the approximated swaption price. Section 9.3 describes the approximate exercise set defined by the logarithm of the ZCBs portfolio geometric mean. Section 9.4 shows the results of numerical tests. Conclusive remarks are presented in last section.

9.2 A Lower Bound on Swaption Prices

In this section, we derive a general formula for lower bound on swaption prices. The price formula (9.1), once a change of measure to the T-forward measure is used, becomes

$$\begin{aligned} C(t, T, \{T_h\}_{h=1}^n, R) &= P(t,T) \cdot \mathbb{E}_t^T \left[\left(\sum_{h=1}^n w_h P(T, T_h) - 1 \right)^+ \right] \\ &= P(t,T) \cdot \mathbb{E}_t^T \left[\left(\sum_{h=1}^n w_h P(T, T_h) - 1 \right) I(\mathcal{A}) \right], \end{aligned} \tag{9.3}$$

where I is the indicator function and $\mathcal{A}$ is the exercise region seen as a subset of the space events Ω

$$\mathcal{A} = \{\omega \in \Omega : \sum_{h=1}^n w_h P(T, T_h) \geq 1\}.$$

Indeed, we observe that for any event set $\mathcal{G} \subset \Omega$

$$\begin{aligned} &\mathbb{E}_t^T \left[\left(\sum_{h=1}^n w_h P(T, T_h) - 1 \right)^+ \right] \geq \mathbb{E}_t^T \left[\left(\sum_{h=1}^n w_h P(T, T_h) - 1 \right)^+ I(\mathcal{G}) \right] \\ &\geq \mathbb{E}_t^T \left[\left(\sum_{h=1}^n w_h P(T, T_h) - 1 \right) I(\mathcal{G}) \right]. \end{aligned}$$

Then by discounting we obtain

$$C(t, T, \{T_h\}_{h=1}^n, R) \geq LB(\mathcal{G}) := P(t,T) \cdot \mathbb{E}_t^T \left[\left(\sum_{h=1}^n w_h P(T, T_h) - 1 \right) I(\mathcal{G}) \right], \tag{9.4}$$

i.e. $LB(\mathcal{G})$ is a lower bound to the swaption price for all possible sets $\mathcal{G}$.

9.2.1 Affine Models

For affine interest rate models the price at T of a zero coupon bond with expiration date T_h can be written as the exponential of a linear combination of the state variables

$$P(T, T_h) = e^{\sum_{j=1}^d b_{h,j} X_j(T) + a_h} = e^{\mathbf{b}_h^\top \mathbf{X}(T) + a_h}, \tag{9.5}$$

where $a_h = A(T_h - T)$ and $\mathbf{b}_h = \mathbf{B}(T_h - T)$ are functions of the payment date T_h and are typical of each model. We know from Duffie and Kan (1996) and Duffie et al. (2000) that under certain regularity conditions, the functions $A(\tau)$ and $\mathbf{B}(\tau)$ are the solution of a system of $d + 1$ ordinary differential equations that are completely determined by the specification of the risk-neutral dynamics of the short rate. These equations can be solved through numerical integration starting from the initial conditions $A(0) = 0, \mathbf{B}(0) = \underline{0}$ and the solutions are known in closed form for most common models.

Moreover from Duffie et al. (2000), we know that for affine models the risk neutral expected value of an exponential payoff has the form

$$\mathbb{E}_t\left[e^{-\int_t^T r(\mathbf{X}(u))du} e^{i\boldsymbol{\lambda}^\top \mathbf{X}(s)}\right] = e^{\tilde{A}(s-t,T-s,\boldsymbol{\lambda})+\tilde{\mathbf{B}}(s-t,T-s,\boldsymbol{\lambda})^\top \mathbf{X}(t)},$$

where $\mathbf{X}$ and $\boldsymbol{\lambda}$ are in $\mathbb{R}^d$ and the functions $\tilde{A}(\tau, T-s, \boldsymbol{\lambda})$ and $\tilde{\mathbf{B}}(\tau, T-s, \boldsymbol{\lambda})$ are solutions of the same ODE system of the zero coupon bond case, but with different initial conditions ($\tilde{A}(0, T-s, \boldsymbol{\lambda}) = A(T-s)$ and $\tilde{\mathbf{B}}(0, T-s, \boldsymbol{\lambda}) = i\boldsymbol{\lambda} + \mathbf{B}(T-s)$). Then the T-forward characteristic function of the model factors $\mathbf{X}$ can be obtained by

$$\begin{aligned}\mathbb{E}_t^T\left[e^{i\boldsymbol{\lambda}^\top \mathbf{X}(s)}\right] &= \frac{1}{P(t,T)}\mathbb{E}_t\left[e^{-\int_t^T r(\mathbf{X}(u))du} e^{i\boldsymbol{\lambda}^\top \mathbf{X}(s)}\right] \\ &= e^{\tilde{A}(s-t,T-s,\boldsymbol{\lambda})-A(T-t)+(\tilde{\mathbf{B}}(s-t,T-t,\boldsymbol{\lambda})^\top-\mathbf{B}(T-t)^\top)\mathbf{X}(t)}.\end{aligned}$$

Since we are interested in the case $s = T$ (forward measure at expiry date of the option), then we define the function

$$\Phi(\boldsymbol{\lambda}) = \mathbb{E}_t^T\left[e^{i\boldsymbol{\lambda}^\top \mathbf{X}(T)}\right] = e^{\tilde{A}(T-t,\boldsymbol{\lambda})-A(T-t)+(\tilde{\mathbf{B}}(T-t,\boldsymbol{\lambda})^\top-\mathbf{B}(T-t)^\top)\mathbf{X}(t)}, \tag{9.6}$$

where $\tilde{A}(\tau, \boldsymbol{\lambda}) = \tilde{A}(\tau, 0, \boldsymbol{\lambda})$ and $\tilde{\mathbf{B}}(\tau, \boldsymbol{\lambda}) = \tilde{\mathbf{B}}(\tau, 0, \boldsymbol{\lambda})$.

We define the set $\mathcal{G}$ using a linear function of the state variates

$$\mathcal{G} = \{\omega \in \Omega : g(\mathbf{X}(T)) \geq k\} = \{\omega \in \Omega : \boldsymbol{\beta}^\top \mathbf{X}(T) + \alpha \geq k\},$$

where $\boldsymbol{\beta}$ is a constant vector of length d, $\alpha \in \mathbb{R}$ and k is a free parameter. k can be chosen such that it optimizes the value of the lower bound.

Proposition 1 *The lower bound to the European swaption price, for affine interest rate models, is given by the following formula*

$$\widehat{LB}(t, T, \{T_h\}_{h=1}^n, R) = \max_{k\in\mathbb{R}} LB(k; t, T, \{T_h\}_{h=1}^n, R), \tag{9.7}$$

where

$$LB(k; t, T, \{T_h\}_{h=1}^n, R) = P(t, T)\,\frac{e^{-\delta k}}{\pi}\int_0^{+\infty} e^{-i\gamma k}\psi_\delta(\gamma)d\gamma, \tag{9.8}$$

and

$$\psi_\delta(\gamma) = \left(\sum_{h=1}^n w_h e^{a_h}\Phi\left(-i\mathbf{b}_h + (\gamma - i\delta)\boldsymbol{\beta}\right) - \Phi\left((\gamma - i\delta)\boldsymbol{\beta}\right)\right)\frac{e^{(i\gamma+\delta)\alpha}}{i\gamma + \delta}, \tag{9.9}$$

where δ is a positive constant.

Proof See Appendix A.1.

9.3 The Geometric Average Approximate Exercise Region

The approximate exercise set is defined through the logarithm of the geometric average of the portfolio of zero coupon bonds

$$\begin{aligned}
\mathcal{G} &= \{\omega \in \Omega : g(\mathbf{X}(T)) \geq k\},\\
G(\mathbf{X}(T)) &= \prod_{h=1}^n P(T, T_h)^{w_h},\\
g(\mathbf{X}(T)) &= \ln\left(G(\mathbf{X}(T))\right) = \sum_{h=1}^n w_h \ln(P(T, T_h)).
\end{aligned}$$

In particular $\mathcal{G}$ and $g(\mathbf{X})$ are given by

$$\mathcal{G} = \{\omega \in \Omega : \boldsymbol{\beta}^\top \mathbf{X}(T) + \alpha \geq k\},$$

where $P(T, T_h) = e^{\mathbf{b}_h{}^\top \mathbf{X}(T) + a_h}$, $\boldsymbol{\beta} = \sum_{h=1}^n w_h \mathbf{b}_h$ and $\alpha = \sum_{h=1}^n w_h a_h$.

Since we don't know the optimum value of the parameter k, then the pricing method requires the maximization of the lower bound, $LB(k; t, T, \{T_h\}_{h=1}^n, R)$, seen as a function of k.

The optimization can be accelerated looking for a good starting point. We suggest the following

$$\tilde{k} = \log\left(\frac{1}{\sum_{h=1}^n w_h}\right) = -\log\left(\sum_{h=1}^n R(T_h - T_{h-1})\right).$$

According to this choice $\mathcal{G}_{\tilde{k}} = \{\omega \in \Omega : g(\mathbf{X}(T)) \geq \tilde{k}\}$ is the greatest possible subset of the true exercise region, $\mathcal{A}$.

In fact normalizing the weights, the expression of the true exercise region can be rewritten as

$$\begin{aligned}\mathcal{A} &= \{\omega \in \Omega : \sum_{h=1}^{n} w_h P(T, T_h) \geq 1\} = \{\omega \in \Omega : \sum_{h=1}^{n} \tilde{w}_h P(T, T_h) \geq e^{\tilde{k}}\} \\ &= \{\omega \in \Omega : A(\mathbf{X}) \geq e^{\tilde{k}}\},\end{aligned}$$

where $A(\mathbf{X})$ is the arithmetic mean of the ZCBs portfolio, $\tilde{w}_h = \frac{w_h}{\sum_{h=1}^{n} w_h}$ and so $\sum_{h=1}^{n} \tilde{w}_h = 1$.

By the arithmetic-geometric inequality we know that $A(\mathbf{X}) \geq G(\mathbf{X})\ \forall \mathbf{X}$, then $\forall k > \tilde{k}$

$$\mathcal{A} \supseteq \mathcal{G}_{\tilde{k}} \supseteq \mathcal{G}_k.$$

Instead if $k < \tilde{k}$ then it is no more guaranteed that $\mathcal{G}_k$ is a subset of the true exercise region.

9.4 Models and Numerical Results

This section presents the examined models and discuss the numerical results.

9.4.1 Affine Gaussian Models

Affine Gaussian models assign the following stochastic differential equation (S.D.E.) to the state variable $\mathbf{X}$

$$\begin{aligned} d\mathbf{X}(t) &= K(\boldsymbol{\theta} - \mathbf{X}(t))\, dt + \Sigma\, d\mathbf{W}(t), \\ \mathbf{X}(0) &= \mathbf{x}_0, \end{aligned}$$

where $\mathbf{W}_t$ is a standard d-dimensional Brownian motion, K is a $d \times d$ diagonal matrix and Σ is a $d \times d$ triangular matrix. The short rate is obtained as a linear combination of the state vector $\mathbf{X}$; it is always possible to rescale the components $X_i(t)$ and assume that

$$r(t) = \phi + \sum_{i=1}^{d} X_i(t)$$

without loss of generality ($\phi \in \mathbb{R}$).

The T-forward characteristic function of $\mathbf{X}$ is

$$\Phi(\boldsymbol{\lambda}) = \mathbb{E}_t^T\left[e^{i\boldsymbol{\lambda}^\top \mathbf{X}(T)}\right] = e^{i\boldsymbol{\lambda}^\top \boldsymbol{\mu}(t,T)-\frac{1}{2}\boldsymbol{\lambda}^\top V(t,T)\boldsymbol{\lambda}},$$

where $\boldsymbol{\mu}$ is the T-forward expected value and V is the covariance matrix (assuming K is diagonal):

$$\begin{aligned}
\boldsymbol{\mu}(t,s) &= \mathbb{E}_t^T[\mathbf{X}(s)] = \mathbb{E}_t[\mathbf{X}(s)] - (I_{n\times n} - e^{-K(s-t)})K^{-1}\Sigma\Sigma^\top(K^{-1})^\top \\
&\qquad \mathbf{g} - Ve^{-K^\top(T-s)}(K^{-1})^\top\mathbf{g} \\
\mathbb{E}_t[\mathbf{X}(s)] &= e^{-K(s-t)}(X(t)-\boldsymbol{\theta}) + \boldsymbol{\theta} \\
V_{ij}(t,s) &= (\Sigma\Sigma^\top)_{ij}\left(\frac{1-e^{-(K_{ii}+K_{jj})(s-t)}}{K_{ii}+K_{jj}}\right).
\end{aligned}$$

where $\mathbf{g} = [1,1,1,\ldots,1]^\top$ is a column vector of length d.

For this type of process, the lower bound can be calculated analytically

$$LB(k;t,T,\{T_h\}_{h=1}^n,R) = P(t,T)\left(\sum_{h=1}^n w_h e^{a_h+\mathbf{b}_h^\top\boldsymbol{\mu}+\frac{1}{2}V_h+\frac{1}{2}d_h^2}N(d_h-d) - N(-d)\right), \tag{9.10}$$

where

$$\begin{aligned}
d &= \frac{k-\boldsymbol{\beta}^\top\boldsymbol{\mu}-\alpha}{\sqrt{\boldsymbol{\beta}^\top V\boldsymbol{\beta}}}, \\
d_h &= \mathbf{b}_h^\top\mathbf{v}, \\
V_h &= \mathbf{b}_h^\top(V-\mathbf{v}\mathbf{v}^\top)\mathbf{b}_h, \\
\mathbf{v} &= \frac{V\boldsymbol{\beta}}{\sqrt{\boldsymbol{\beta}^\top V\boldsymbol{\beta}}},
\end{aligned}$$

and $N(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x e^{-\frac{y^2}{2}}dy$ is the cumulative distribution function of a standard normal.

See details in Appendix A.2.

9.4.2 Multi-factor Cox-Ingersoll-Ross (CIR) Model

The state vector of the model evolves according to the following system of SDE

$$\begin{aligned}
dX_i(t) &= a_i(\theta_i - X_i(t))dt + \sigma_i\sqrt{X_i(t)}dW^i(t), \\
\mathbf{X}(0) &= \mathbf{x}_0,
\end{aligned}$$

where $i = 1, \dots, d$, $W^i(t)$ are independent standard Brownian motions, a_i, θ_i and σ_i are positive constants. The short rate is obtained by

$$r(t) = \phi + \sum_{i=1}^{d} X_i(t),$$

where $\phi \in \mathbb{R}$.

Using the results in Collin-Dufresne and Goldstein (2002), we can calculate the zero coupon bond price and the T-forward characteristic function. In particular we deduce the functions $A(\tau)$, $\mathbf{B}(\tau)$, $\tilde{A}(\tau, \boldsymbol{\lambda})$ and $\tilde{\mathbf{B}}(\tau, \boldsymbol{\lambda})$ in formula (9.6),

$$\begin{aligned}
A(\tau) &= -\phi\, \tau + \sum_{j=1}^{d} \left[\frac{2a_j\theta_i}{h_j - a_j}\tau - \frac{2a_j\theta_j}{\sigma_j^2} \ln\left(\frac{(a_j + h_j)(e^{h_j\tau} - 1) + 2h_j}{2h_j} \right) \right], \\
B_j(\tau) &= \frac{-2\,(e^{h_j\tau} - 1)}{(a_j + h_j)(e^{h_j\tau} - 1) + 2h_j}, \\
\tilde{A}(\tau, \boldsymbol{\lambda}) &= -\phi\, \tau + \sum_{j=1}^{d} \left[\frac{2a_j\theta_i}{h_j - a_j}\tau - \frac{2a_j\theta_j}{\sigma_j^2} \ln\left(\frac{\sigma_j^2((i\lambda_j + \mu_j^+) - (i\lambda + \mu_j^-)e^{h_j\tau})}{2h_j} \right) \right], \\
\tilde{B}_j(\tau, \boldsymbol{\lambda}) &= \frac{i\lambda_j(\mu_j^- - \mu_j^+ e^{h_j\tau}) + \frac{2(e^{h_j\tau}-1)}{\sigma_j^2}}{i\lambda_j(e^{h_j\tau} - 1) + (\mu_j^- e^{h_j\tau} - \mu_j^+)},
\end{aligned}$$

where $h_j = \sqrt{a_j^2 + 2\sigma_j^2}$ and $\mu_j^{\pm} = \frac{-a_j \pm h_j}{\sigma_j^2}$.

9.4.3 Gaussian Model with Double Exponential Jumps

The vector of model factors evolves according to the following S.D.E.

$$\begin{aligned}
d\mathbf{X}(t) &= K(\boldsymbol{\theta} - \mathbf{X}(t))\, dt + \Sigma\, d\mathbf{W}(t) + d\mathbf{Z}^+(t) - d\mathbf{Z}^-(t), \\
\mathbf{X}(0) &= \mathbf{x}_0,
\end{aligned}$$

where $\mathbf{W}_t$ is a standard d-dimensional Brownian motion, K is a $d \times d$ diagonal matrix, Σ is a $d \times d$ triangular matrix and $\mathbf{Z}^{\pm}$ are pure jumps processes whose jumps have fixed probability distribution ν on $\mathbb{R}^d$ and constant intensity $\mu^{\pm}$. The short rate is obtained as a linear combination of the state vector $\mathbf{X}$. In particular $\mathbf{Z}^{\pm}$ are compound Poisson processes with jump size having exponential distribution

$$Z_l^{\pm} = \sum_{j=1}^{N^{\pm}(t)} Y_{j,l}^{\pm},$$

where $l = 1, \dots, d$ is the factor index, $N^{\pm}(t)$ are Poisson process with intensity $\frac{\mu^{\pm}}{d}$[3] and $Y^{\pm}_{j,l}$, for a fixed l, are independent identically distributed exponential random variables of mean parameters $m^{\pm}_l$:

$$Y^{\pm}_{j,l} \sim \nu(m^{\pm}_l) = \frac{1}{m^{\pm}_l} \exp\left(\frac{y}{m^{\pm}_l}\right).$$

Since $\mu^{\pm}$ do not depend on $\mathbf{X}$, we know from Duffie et al. (2000) that: (a) the functions $\mathbf{B}(\tau)$ and $\tilde{\mathbf{B}}(\tau, \boldsymbol{\lambda})$ in formula (9.6) are independent of jumps and they are calculated as in affine Gaussian model case; (b) the functions $A(\tau)$ and $\tilde{A}(\tau, \boldsymbol{\lambda})$ in formula (9.6) are sum of two components, diffusive and jump: $A(\tau) = A^D(\tau) + A^J(\tau)$ and $\tilde{A}(\tau, \boldsymbol{\lambda}) = \tilde{A}^D(\tau, \boldsymbol{\lambda}) + \tilde{A}^J(\tau, \boldsymbol{\lambda})$. The diffusive part can be obtained as in affine Gaussian model case.

Then we obtain that

$$\Phi(\boldsymbol{\lambda}) = \mathbb{E}^T_t\left[e^{i\boldsymbol{\lambda}^\top \mathbf{X}(T)}\right] = \Phi^D(\boldsymbol{\lambda})\, e^{\tilde{A}^J(T-t,\boldsymbol{\lambda}) - A^J(T-t)}, \tag{9.11}$$

where $\Phi^D(\boldsymbol{\lambda})$ is the T-forward characteristic function of affine Gaussian model. The jumps component is calculated using the characteristic function of the jumps size distribution ν

$$\begin{aligned}
&\theta^{\pm}(\mathbf{c}) = \int_{\mathbb{R}^d} e^{\mathbf{c}^\top \mathbf{y}} d\nu(\mathbf{y}) = \frac{1}{d}\sum_{j=1}^{d} \frac{1}{1 - m^{\pm}_j c_j} = \frac{1}{d}\sum_{j=1}^{d} \frac{\eta^{\pm}_j}{\eta^{\pm}_j - c_j},\\
&\tilde{A}^J(T-t, \boldsymbol{\lambda}) = \mu^+ \int_t^T ds\, \left(\theta^+\left(\tilde{\mathbf{B}}(s, \boldsymbol{\lambda})\right) - 1\right) + \mu^- \int_t^T ds\, \left(\theta^-\left(\tilde{\mathbf{B}}(s, \boldsymbol{\lambda})\right) - 1\right),\\
&A^J(T-t) = \tilde{A}^J(T-t, \mathbf{0}),
\end{aligned}$$

where $c \in \mathbb{C}^d$ and $\eta^{\pm}_j = \frac{1}{m^{\pm}_j}$. The function $\tilde{A}^J(\tau, \boldsymbol{\lambda})$ is available in closed form

$$\begin{aligned}
\tilde{A}^J(\tau, \boldsymbol{\lambda}) = {}& \frac{\mu^+}{d} \sum_{j=1}^{d} \frac{-\tau}{1 + \eta^+_j K_{jj}} + \frac{\eta^+_j}{1 + \eta^+_j K_{jj}} \log\left(\frac{(1 + i\lambda_j K_{jj}) e^{-K_{jj}\tau} - 1 - \eta^+_j K_{jj}}{K_{jj}(i\lambda_j - \eta^+_j)}\right)\\
&+ \frac{\mu^-}{d} \sum_{j=1}^{d} \frac{-\tau}{1 - \eta^-_j K_{jj}} - \frac{\eta^-_j}{1 - \eta^-_j K_{jj}} \log\left(\frac{(1 + i\lambda_j K_{jj}) e^{-K_{jj}\tau} - 1 + \eta^-_j K_{jj}}{K_{jj}(i\lambda_j + \eta^-_j)}\right).
\end{aligned}$$

[3] $N^{\pm}(t)$ represent the number of positive or negative jumps before the time t.

9.4.4 Balduzzi, Das, Foresi and Sundaram Model

In the model proposed in Balduzzi et al. (1996), the interest rate follows the same stochastic process as in CIR model, but the long-mean $\theta(t)$ and the variance $V(t)$ are stochastic, according to the following system of SDEs

$$\begin{cases} dr(t) = k\,(\theta(t) - r(t))\,dt + \sqrt{V(t)}\,dW(t) \\ d\theta(t) = \alpha\,(\beta - \theta(t))\,dt + \eta\,dZ(t) \\ dV(t) = a\,(b - V(t))\,dt + \phi\,\sqrt{V(t)}\,dY(t) \\ dW(t)dY(t) = \rho\,dt \\ r(0) = r_0,\ \theta(0) = \theta_0 \text{ and } V(0) = v_0 \end{cases}$$

k, α, β, η, a, b and ϕ are positive constants and $\rho \in [-1, 1]$. In order to align the notation with previous sections we denote $\mathbf{X}(t) = [r(t), \theta(t), V(t)]^\top$. In this model the characteristic function of the state vector $\mathbf{X}$ can not be obtained in closed form,[4] but can be calculated numerically, solving the following system of ordinary differential equations

$$\begin{aligned} \frac{d\tilde{A}(\tau,\boldsymbol{\lambda})}{d\tau} &= -\alpha\beta\tilde{B}_2(\tau,\boldsymbol{\lambda}) + \frac{1}{2}\eta^2\tilde{B}_2(\tau,\boldsymbol{\lambda})^2 - ab\tilde{B}_3(\tau,\boldsymbol{\lambda}), \\ \tilde{B}_1(\tau,\boldsymbol{\lambda}) &= \lambda_1 e^{-k\tau} - \frac{1-e^{-k\tau}}{k}, \\ \tilde{B}_2(\tau,\boldsymbol{\lambda}) &= (k\lambda_1 + 1)\frac{e^{-k\tau} - e^{-\alpha\tau}}{\alpha - k} + \lambda_2 e^{-\alpha\tau} - \frac{1-e^{-\alpha\tau}}{\alpha}, \\ \frac{d\tilde{B}_3(\tau,\boldsymbol{\lambda})}{d\tau} &= \frac{1}{2}\tilde{B}_1(\tau,\boldsymbol{\lambda})^2 + a\tilde{B}_3(\tau,\boldsymbol{\lambda}) + \frac{1}{2}\phi^2\tilde{B}_3(\tau,\boldsymbol{\lambda})^2 + \rho\phi\tilde{B}_1(\tau,\boldsymbol{\lambda})\tilde{B}_3(\tau,\boldsymbol{\lambda}). \end{aligned}$$

The functions $A(\tau)$ and $\mathbf{B}(\tau)$ (i.e. the zero coupon bond price) can be obtained solving the previous system and setting $\boldsymbol{\lambda} = \underline{0}$.

9.4.5 Numerical Results

Apart from the Vasicek model for which a simple closed form solution is available, Monte Carlo is used as benchmark for the computation of the true swaption price. The 97.5% mean-centered Monte Carlo Confidence Interval[5] is used as measure of

[4] A semi-analytical solution for functions $\tilde{A}(\tau,\boldsymbol{\lambda})$ and $\tilde{B}_3(\tau,\boldsymbol{\lambda})$ is available but it requires the evaluation of Kummer's functions of the first and second kind. Kummer's functions are not analytic but have series and integral representation. However we find that the numerical solution of the ODE system is much more efficient than the evaluation of the semi analytical form.

[5] Note that we use the quantile function of a Student's t distribution.

the accuracy of the Monte Carlo method. Simulation of the affine 3-factor Gaussian and 2-factor CIR models is implemented by sampling from the exact distribution. The Gaussian with jumps model and the BDFS model are simulated using an Euler-Maruyama scheme with a time step of 0.0005. The number of simulations is chosen depending on the complexity of the model and it is specified in each table caption. Antithetic variates technique is also used for the affine 3-factor Gaussian model and the BDFS model.

For the Vasicek model and the 3-factor Gaussian model, lower bound is obtained via the closed formula described in Sect. 9.4.1. For the 2-factor CIR model and the Gaussian with jumps model, the integrals involved in the lower bound are evaluated by a Gauss-Kronrod quadrature rule, using Matlab's built-in function **quadgk**. For the BDFS model the system of ordinary differential equations is solved numerically using the Matlab function **ode45** based on the Dormand–Prince method. Due to the complexity of the problem we adopt a Gauss-Legendre quadrature rule. The calculus of the lower bound with a geometric mean region $\mathcal{G}$ involves the optimization of the function $LB(k; t, T, \{T_h\}_{h=1}^n, R)$ with respect to the parameter k. The optimization is performed via Matlab functions **fminunc**.

Another important fact is that our lower bound formula is very suitable to be used as a control variate to reduce Monte Carlo error. The approximated formula is easy to implement in a Monte Carlo scheme and turns out to be very effective (see Caldana et al. (2014) for details).

Swaption prices for different tenor and maturities are reported in Tables 9.1, 9.2, 9.3, 9.4, and 9.5 with the relative overall computing time for each pricing method.

Moreover, Figs. 9.1, 9.2, 9.3, 9.4, and 9.5 compare graphically the relative error of the three proposed pricing methods.

9.4.5.1 Vasicek Model, Three-Factors Gaussian Model and Cox-Ingersoll and Ross Model

We verify the accuracy of our new lower bound using models and parameter values already examined in literature[6]

- Vasicek model: $K = 0.05$, $\theta = 0.05$, $\Sigma = 0.01$, $x_0 = 0.05$ and $\phi = 0$;
- 3-factors Gaussian model: $K = \begin{bmatrix} 1.0 & 0 & 0 \\ 0 & 0.2 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}$, $\theta = [0,\ 0,\ 0]^\top$, $\boldsymbol{\sigma} = [0.01,\ 0.005,\ 0.002]^\top$, $\rho = \begin{bmatrix} 1 & -0.2 & -0.1 \\ -0.2 & 1 & 0.3 \\ -0.1 & 0.3 & 1 \end{bmatrix}$, $\Sigma = \mathrm{diag}(\boldsymbol{\sigma}) \cdot \mathrm{chol}(\rho)$,[7] $x_0 = [0.01,\ 0.005,\ -0.02]$ and $\phi = 0.06$;

[6] Schrager and Pelsser (2006) and Duffie and Singleton (1997) for the 2-factors C.I.R. model.

[7] $\mathrm{diag}(\boldsymbol{\sigma})$ means the diagonalization of the vector $\boldsymbol{\sigma}$ and $\mathrm{chol}(\rho)$ means the Cholesky decomposition of the correlation matrix ρ, where $\boldsymbol{\sigma}$ and ρ are the volatility vector and the correlation matrix, respectively, of the original paper.

Table 9.1 The three tables represent matrices of swaption prices at three different strikes for the Vasicek model. For each swaption we report on the first line the price in basis points and on the second line the error in basis points. The error of the lower bound is the difference from the exact Jamshidian method (Jam.)

Vasicek model						
Payer swaptions (ATMF)						
Option mat.	1		2		5	
Swap length	LB	Jam.	LB	Jam.	LB	Jam.
1	35.670 10^{-4}	35.670	46.836 10^{-4}	46.836	59.501 10^{-4}	59.501
2	67.953 10^{-4}	67.953	89.234 10^{-4}	89.234	113.412 10^{-4}	113.412
5	147.645 10^{-4}	147.645	193.957 10^{-4}	193.957	246.875 10^{-4}	246.875
10	238.273 10^{-4}	238.273	313.243 10^{-4}	313.243	399.674 10^{-4}	399.674
Payer swaptions (ITMF: 0.85 × ATMF)						
Option mat.	1		2		5	
Swap length	LB	Jam.	LB	Jam.	LB	Jam.
1	80.591 10^{-4}	80.591	86.861 10^{-4}	86.861	91.405 10^{-4}	91.405
2	155.872 10^{-4}	155.872	167.452 10^{-4}	167.452	175.622 10^{-4}	175.622
5	353.282 10^{-4}	353.282	376.199 10^{-4}	376.199	391.032 10^{-4}	391.032
10	605.661 10^{-4}	605.661	637.301 10^{-4}	637.301	654.439 10^{-4}	654.439
Payer swaptions (OTMF: 1.15 × ATMF)						
Option mat.	1		2		5	
Swap length	LB	Jam.	LB	Jam.	LB	Jam.
1	11.247 10^{-4}	11.247	21.169 10^{-4}	21.169	35.825 10^{-4}	35.825
2	20.781 10^{-4}	20.781	39.551 10^{-4}	39.551	67.525 10^{-4}	67.525
5	41.394 10^{-4}	41.394	81.330 10^{-4}	81.330	142.400 10^{-4}	142.400
10	58.744 10^{-4}	58.744	121.059 10^{-4}	121.059	220.008 10^{-4}	220.008

- 2-factors Cox-Ingersoll and Ross model: $\mathbf{a} = [0.5080,\ -0.0010]^\top$, $\boldsymbol{\theta} = [0.4005,\ -0.7740]^\top$, $\boldsymbol{\sigma} = [0.023,\ 0.019]^\top$, $x_0 = [0.374,\ 0.258]$ and $\phi = -0.58$.

  Numerical results for these models are shown in Tables 9.1, 9.2 and 9.3.

Table 9.2 The three tables represent matrices of swaption prices at three different strikes for the 3-factors Gaussian model. The bottom line of each table provides the overall computation time for the different pricing methods. For each swaption we report the price in basis points estimated with the Monte Carlo method, MC, the Lower Bound approximation (LB) and the Monte Carlo method with the control variable technique, MC (CV). Monte Carlo prices without and with control variable method are estimated using 10^7 and respectively, 10^5 simulations, antithetic variates method and the exact probability distribution of the state variables at the maturity date of the swaption. Below each Monte Carlo price, the confidence interval at 97.5% is reported in basis point

Three-factor Gaussian model									
Payer swaptions (ATMF)									
Opt. mat.	1			2			5		
Swap length	MC	LB	MC (CV)	MC	LB	MC (CV)	MC	LB	MC (CV)
1	20.814 0.010	20.817	20.817 10^{-4}	23.554 0.011	23.554	23.554 10^{-4}	23.217 0.011	23.207	23.207 10^{-4}
2	33.114 0.015	33.119	33.119 10^{-4}	38.430 0.018	38.434	38.434 10^{-4}	38.728 0.018	38.722	38.722 10^{-4}
5	53.325 0.025	53.312	53.312 10^{-4}	63.676 0.029	63.686	63.686 10^{-4}	65.673 0.030	65.683	65.683 10^{-4}
10	65.583 0.030	65.579	65.583 0.001	79.090 0.036	79.062	79.067 0.001	82.164 0.037	82.156	82.159 10^{-4}
Overall time (sec)	MC 32.045	LB 0.122	MC (CV) 3.657						
Payer swaptions (ITMF: 0.85 × ATMF)									
Opt. mat.	1			2			5		
Swap length	MC	LB	MC (CV)	MC	LB	MC (CV)	MC	LB	MC (CV)
1	79.446 0.003	79.445	79.445 10^{-4}	78.404 0.005	78.404	78.404 10^{-4}	69.446 0.005	69.442	69.442 10^{-4}
2	154.563 0.003	154.563	154.563 10^{-4}	150.909 0.005	150.911	150.911 10^{-4}	131.95 0.007	131.949	131.949 10^{-4}
5	361.470 0.001	361.469	361.469 10^{-4}	346.275 0.003	346.275	346.275 10^{-4}	295.162 0.006	295.162	295.162 10^{-4}
10	636.982 0.001	636.982	636.982 10^{-4}	604.809 0.002	604.81	604.81 10^{-4}	508.838 0.003	508.840	508.840 10^{-4}
Overall time (sec)	MC 32.023	LB 0.117	MC (CV) 3.543						
Payer swaptions (OTMF: 1.15 × ATMF)									
Opt. mat.	1			2			5		
Swap length	MC	LB	MC (CV)	MC	LB	MC (CV)	MC	LB	MC (CV)
1	1.571 0.003	1.570	1.570 10^{-4}	2.823 0.005	2.824	2.824 10^{-4}	3.798 0.006	3.794	3.794 10^{-4}
2	1.065 0.003	1.065	1.065 10^{-4}	2.611 0.006	2.612	2.612 10^{-4}	4.324 0.008	4.322	4.322 10^{-4}
5	0.150 0.001	0.150	0.150 10^{-4}	0.904 0.004	0.905	0.905 10^{-4}	2.569 0.007	2.569	2.57 10^{-4}
10	0.003 10^{-4}	0.003	0.003 10^{-4}	0.076 0.001	0.076	0.076 10^{-4}	0.517 0.003	0.516	0.517 10^{-4}
Overall time (sec)	MC 32.024	LB 0.113	MC (CV) 3.541						

Table 9.3 The three tables represent matrices of swaption prices at three different strikes for the 2-factor C.I.R. model. The bottom line of each table provides the overall computation time for the different pricing methods. For each swaption we report the price in basis points estimated with the Monte Carlo method, MC, the Lower Bound approximation (LB) and the Monte Carlo method with the control variable technique, MC (CV). Monte Carlo prices without and with control variable method are estimated using 10^7 and respectively, 10^5 simulations and the exact probability distribution of the state variables at the maturity date of the swaption. Below each Monte Carlo price, the confidence interval at 97.5% is reported in basis point

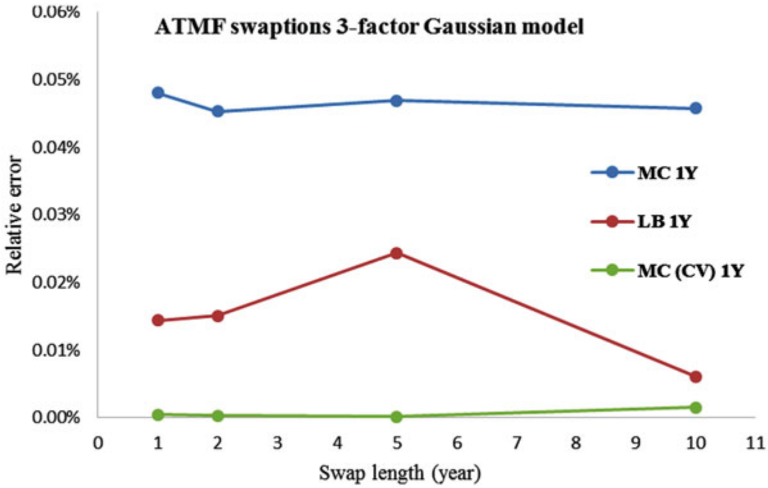
Two-factor Cox-Ingersoll-Ross model									
Payer swaptions (ATMF)									
Opt. mat.	1			2			5		
Swap length	MC	LB	MC (CV)	MC	LB	MC (CV)	MC	LB	MC (CV)
1	48.490 0.044	48.466	48.466 10^{-4}	59.386 0.054	59.361	59.361 10^{-4}	67.006 0.060	66.970	66.970 10^{-4}
2	85.908 0.078	85.871	85.871 10^{-4}	106.938 0.098	106.890	106.890 10^{-4}	123.894 0.113	123.830	123.830 10^{-4}
5	169.474 0.157	169.427	169.428 10^{-4}	216.993 0.202	216.879	216.881 10^{-4}	261.486 0.246	261.366	261.368 10^{-4}
10	265.862 0.251	265.779	265.818 0.004	345.184 0.331	344.949	344.992 0.005	423.059 0.418	422.857	422.887 0.004
Overall time (sec)	MC 23.118	LB 1.456	MC (CV) 1.924						
Payer swaptions (ITMF: 0.85 × ATMF)									
Opt. mat.	1			2			5		
Swap length	MC	LB	MC (CV)	MC	LB	MC (CV)	MC	LB	MC (CV)
1	107.584 0.025	107.577	107.577 10^{-4}	116.846 0.034	116.839	116.839 10^{-4}	114.947 0.043	114.930	114.930 10^{-4}
2	208.047 0.040	208.037	208.037 10^{-4}	222.373 0.058	222.363	222.363 10^{-4}	217.236 0.078	217.208	217.208 10^{-4}
5	475.686 0.065	475.668	475.669 10^{-4}	493.334 0.109	493.301	493.301 10^{-4}	473.362 0.165	473.330	473.331 10^{-4}
10	812.501 0.092	812.470	812.482 0.002	825.291 0.168	825.202	825.219 0.003	778.599 0.275	778.559	778.573 0.002
Overall time (sec)	MC 23.121	LB 1.199	MC (CV) 1.667						
Payer swaptions (OTMF: 1.15 × ATMF)									
Opt. mat.	1			2			5		
Swap length	MC	LB	MC (CV)	MC	LB	MC (CV)	MC	LB	MC (CV)
1	16.004 0.062	15.973	15.973 10^{-4}	24.476 0.072	24.446	24.446 10^{-4}	34.601 0.077	34.546	34.546 10^{-4}
2	23.777 0.113	23.724	23.724 10^{-4}	40.022 0.134	39.964	39.964 10^{-4}	61.943 0.146	61.838	61.838 10^{-4}
5	33.668 0.239	33.565	33.567 10^{-4}	68.868 0.288	68.740	68.742 0.001	124.627 0.323	124.396	124.399 0.001
10	42.602 0.394	42.425	42.459 0.005	99.260 0.483	99.004	99.045 0.005	196.625 0.555	196.230	196.266 0.005
Overall time (sec)	MC 23.121	LB 1.314	MC (CV) 1.782						

Table 9.4 The three tables represent matrices of swaption prices at three different strikes for the 2-factor Gaussian model with Double Exponential Jumps. The bottom line of each table provides the overall computation time for the different pricing methods. For each swaption we report the price in basis points estimated with the Monte Carlo method, MC, the Lower Bound approximation (LB) and the Monte Carlo method with the control variable technique, MC (CV). Monte Carlo (MC) price is estimated using four millions simulations, an Euler scheme with a time step equal to 0.0025 and the antithetic variates technique. Monte Carlo is also performed using 10^5 simulations and control variates method. Below each Monte Carlo price, the confidence interval at 97.5% is reported in basis point

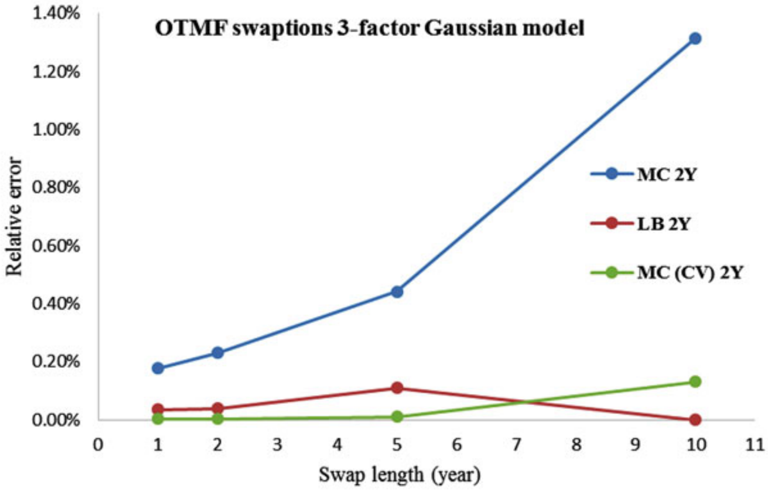
Two-factor Gaussian model with Double Exponential Jumps									
Payer swaptions (ATMF)									
Opt. mat.	1			2			5		
Swap length	MC	LB	MC (CV)	MC	LB	MC (CV)	MC	LB	MC (CV)
1	44.013 0.063	44.046	44.046 10^{-4}	52.773 0.074	52.766	52.766 10^{-4}	59.145 0.082	59.204	59.204 10^{-4}
2	82.968 0.117	83.042	83.042 10^{-4}	102.255 0.143	102.242	102.242 10^{-4}	119.461 0.165	119.629	119.629 10^{-4}
5	96.950 0.137	97.030	97.031 10^{-4}	120.664 0.169	120.652	120.652 10^{-4}	142.807 0.196	143.021	143.021 10^{-4}
10	97.542 0.138	97.623	97.623 10^{-4}	121.396 0.170	121.384	121.384 10^{-4}	143.643 0.198	143.858	143.858 10^{-4}
Overall time (sec)	MC 5.396×10^3	LB 2.643	MC (CV) 84.485						
Payer swaptions (ITMF: 0.85 × ATMF)									
Opt. mat.	1			2			5		
Swap length	MC	LB	MC (CV)	MC	LB	MC (CV)	MC	LB	MC (CV)
1	63.821 0.074	63.878	63.878 10^{-4}	68.253 0.084	68.267	68.267 10^{-4}	69.103 0.088	69.164	69.164 10^{-4}
2	138.588 0.146	138.729	138.729 10^{-4}	148.623 0.169	148.667	148.667 10^{-4}	153.982 0.184	154.172	154.172 10^{-4}
5	349.841 0.214	350.150	350.150 10^{-4}	347.712 0.250	347.888	347.888 10^{-4}	340.988 0.275	341.320	341.320 10^{-4}
10	904.579 0.215	904.933	904.934 10^{-4}	887.761 0.264	887.995	887.995 10^{-4}	856.180 0.305	856.559	856.559 10^{-4}
Overall time (sec)	MC 5.396×10^3	LB 2.643	MC (CV) 84.057						
Payer swaptions (OTMF: 1.15 × ATMF)									
Opt. mat.	1			2			5		
Swap length	MC	LB	MC (CV)	MC	LB	MC (CV)	MC	LB	MC (CV)
1	28.669 0.051	28.679	28.679 10^{-4}	39.757 0.065	39.729	39.729 10^{-4}	50.158 0.076	50.215	50.215 10^{-4}
2	44.305 0.087	44.319	44.319 10^{-4}	66.449 0.117	66.396	66.396 10^{-4}	90.386 0.144	90.535	90.535 10^{-4}
5	9.461 0.041	9.449	9.449 10^{-4}	22.651 0.072	22.573	22.573 10^{-4}	41.176 0.105	41.275	41.275 10^{-4}
10	0.051 0.005	0.048	0.048 10^{-4}	0.299 0.009	0.304	0.304 10^{-4}	1.563 0.019	1.564	1.564 10^{-4}
Overall time (sec)	MC 5.396×10^3	LB 2.643	MC (CV) 84.070						

Table 9.5 The three tables represent matrices of swaption prices at three different strikes for the BDFS. The bottom line of each table provides the overall computation time for the different pricing methods. For each swaption we report the price in basis points estimated with the Monte Carlo method, MC, the Lower Bound approximation (LB) and the Monte Carlo method with the control variable technique, MC (CV). Monte Carlo (MC) price is estimated using four millions simulations, an Euler scheme with a time step equal to 0.0005 and the antithetic variates technique. Monte Carlo is also performed using 10^5 simulations and control variates method. Below each Monte Carlo price, the confidence interval at 97.5% is reported in basis point

Balduzzi, Das, Foresi and Sundaram model									
Payer swaptions (ATMF)									
Opt. mat.	1			2			5		
Swap length	MC	LB	MC (CV)	MC	LB	MC (CV)	MC	LB	MC (CV)
1	85.08	85.00	85.00	102.24	102.17	102.16	97.45	97.45	97.45
	0.06		10^{-3}	0.07		10^{-3}	0.06		10^{-3}
2	148.39	148.27	148.26	177.40	177.28	177.27	167.37	167.38	167.38
	0.10		10^{-3}	0.11		10^{-3}	0.09		10^{-3}
5	250.22	250.01	250.01	296.89	296.68	296.68	276.29	276.28	276.28
	0.17		10^{-3}	0.18		10^{-3}	0.15		10^{-3}
10	290.96	290.71	290.72	345.34	345.10	345.10	321.69	321.68	321.68
	0.19		10^{-3}	0.21		10^{-3}	0.17		10^{-3}
Overall	MC	LB	MC (CV)						
time (sec)	12.256×10^3	894	1.139×10^3						
Payer swaptions (ITMF: 0.85 × ATMF)									
Opt. mat.	1			2			5		
Swap length	MC	LB	MC (CV)	MC	LB	MC (CV)	MC	LB	MC (CV)
1	122.53	122.44	122.44	148.64	148.57	148.57	152.26	152.25	152.25
	0.05		10^{-3}	0.06		10^{-3}	0.04		10^{-3}
2	234.64	234.51	234.50	277.64	277.52	277.51	277.32	277.31	277.31
	0.09		10^{-3}	0.1		10^{-3}	0.07		10^{-3}
5	521.94	521.72	521.72	578.32	578.11	578.11	543.71	543.68	543.68
	0.11		10^{-3}	0.12		10^{-3}	0.08	0	0
10	881.46	881.21	881.21	908.49	908.26	908.26	799.72	799.70	799.70
	0.07		10^{-3}	0.08		10^{-3}	0.07		10^{-3}
Overall	MC	LB	MC (CV)						
time (sec)	12.256×10^3	930	1.176×10^3						
Payer swaptions (OTMF: 1.15 × ATMF)									
Opt. mat.	1			2			5		
Swap length	MC	LB	MC (CV)	MC	LB	MC (CV)	MC	LB	MC (CV)
1	55.91	55.84	55.84	66.35	66.29	66.28	57.30	57.30	57.30
	0.06		10^{-3}	0.07		10^{-3}	0.06		10^{-3}
2	85.74	85.65	85.64	104.05	103.95	103.94	90.40	90.40	90.40
	0.09		10^{-3}	0.11		10^{-3}	0.09		10^{-3}
5	92.5	92.38	92.38	124.12	123.99	123.98	113.7	113.69	113.69
	0.13		10^{-3}	0.16		10^{-3}	0.13		10^{-3}
10	48.09	48.00	48.01	81.30	81.18	81.17	86.16	86.15	86.15
	0.10		10^{-3}	0.14		10^{-3}	0.13		10^{-3}
Overall	MC	LB	MC (CV)						
time (sec)	12.256×10^3	928	1.172×10^3						



(a)



(b)

Fig. 9.1 Three-factor Gaussian model: relative error in percentage. The error of the lower bound is the difference from Monte Carlo value. (**a**) ATMF-1 year maturity. (**b**) OTMF-2 years maturity

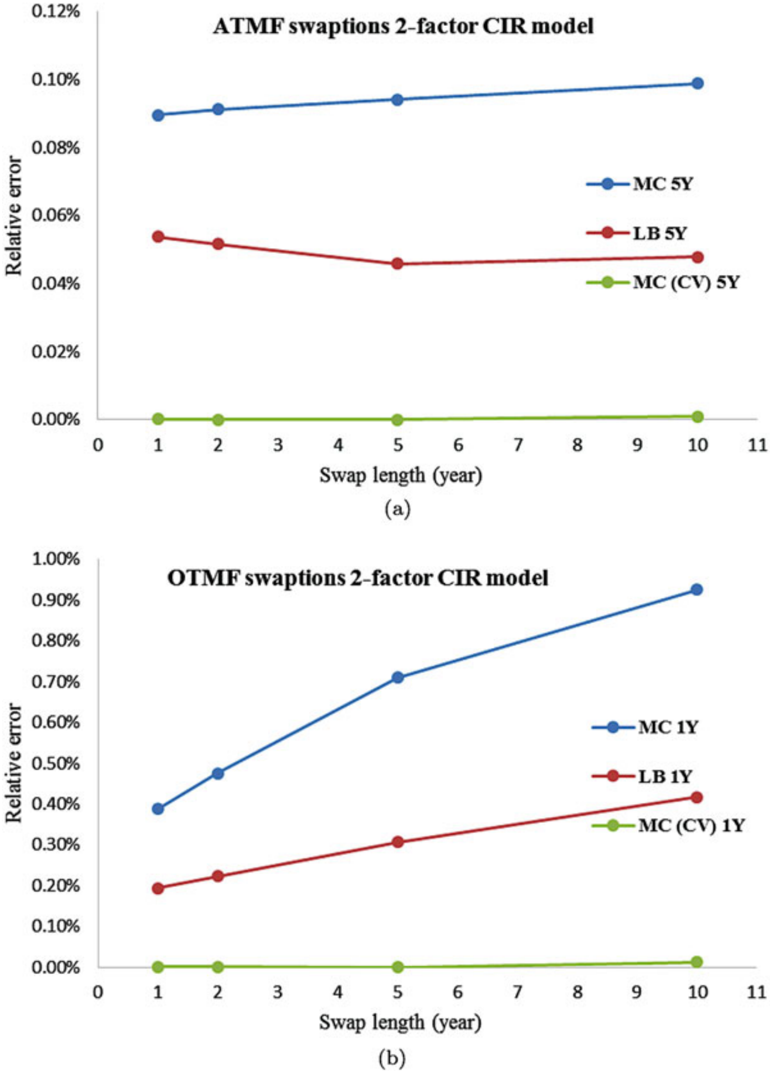


Fig. 9.2 Two-factor Cox-Ingersoll-Ross model: relative error in percentage. The error of the lower bound is the difference from Monte Carlo value. (**a**) ATMF-5 year maturity. (**b**) OTMF-1 years maturity

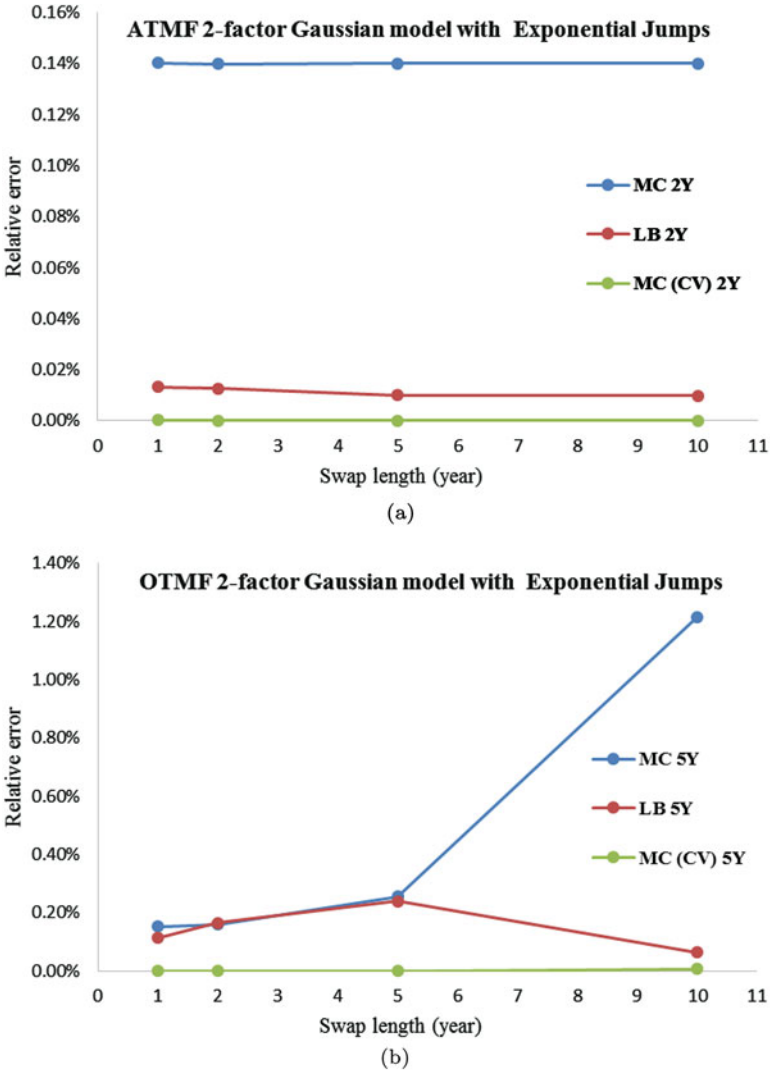
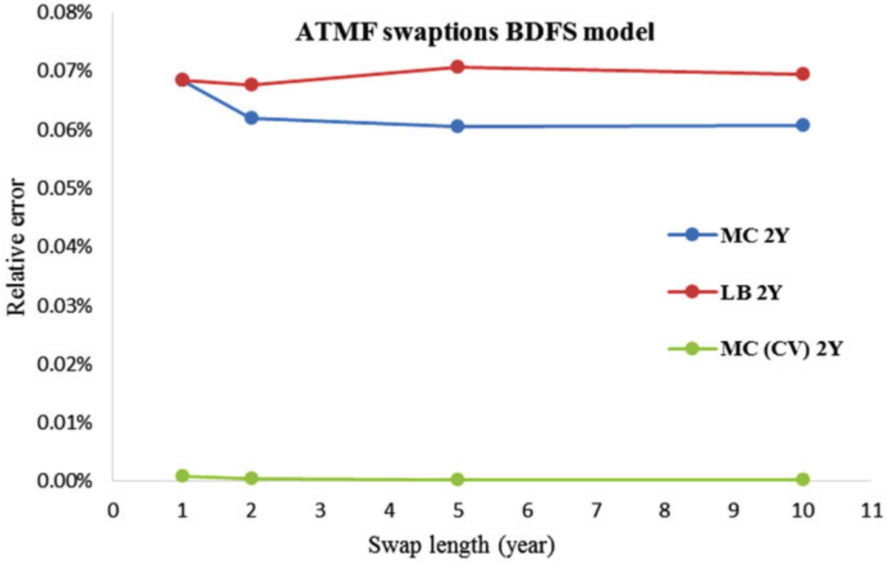
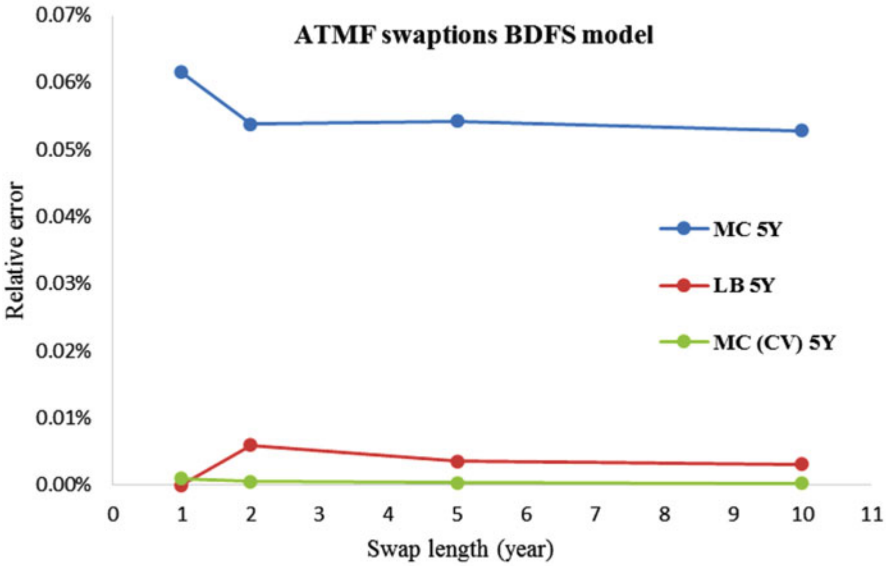


Fig. 9.3 Two-factor Gaussian model with Double Exponential Jumps: relative error in percentage. The error of the lower bound is the difference from Monte Carlo value. (**a**) ATMF-2 year maturity. (**b**) OTMF-5 years maturity



(a)



(b)

Fig. 9.4 Balduzzi, Das, Foresi and Sundaram model: relative error in percentage. The error of the lower bound is the difference from Monte Carlo value. (**a**) ATMF-2 year maturity. (**b**) ATMF-5 years maturity

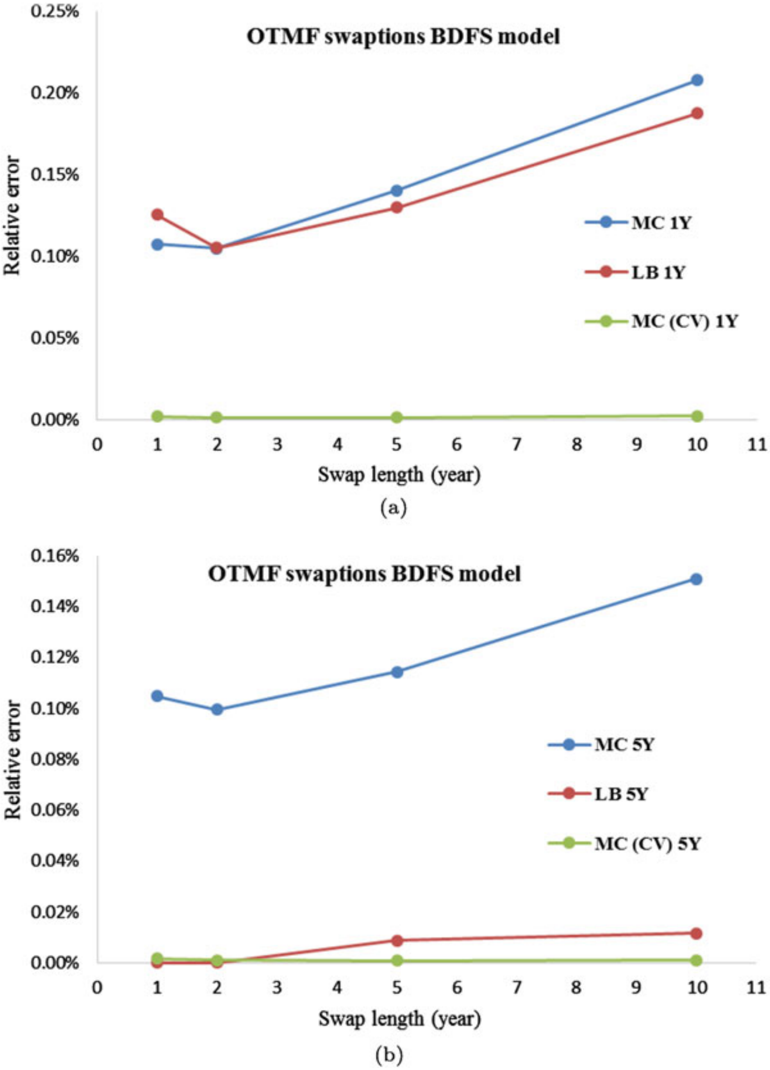


Fig. 9.5 Balduzzi, Das, Foresi and Sundaram model: relative error in percentage. The error of the lower bound is the difference from Monte Carlo value. (**a**) OTMF-2 year maturity. (**b**) OTMF-5 years maturity

9.4.5.2 Two-Factor Gaussian Model with Double Exponential Jumps

We test the affine Gaussian with Jumps interest rate model using the following parameter values

- Gaussian parameters: $K = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.2 \end{bmatrix}$, $\theta = [0,\ 0]^\top$, $\boldsymbol{\sigma} = [0.01,\ 0.005]^\top$, $\rho = \begin{bmatrix} 1 & -0.2 \\ -0.2 & 1 \end{bmatrix}$, $\Sigma = \mathrm{diag}(\boldsymbol{\sigma}) \cdot \mathrm{chol}(\rho)$, $x_0 = [0.01,\ 0.005]$ and $\phi = 0.005$;
- Jumps parameters: $\boldsymbol{\mu}^+ = 0.001$, $\mathbf{m}^+ = [0.01,\ 0.01]$, $\boldsymbol{\mu}^- = 0.001$, $\mathbf{m}^- = [0.01,\ 0.01]$.

Numerical results for this model are shown in Table 9.4.

9.4.5.3 Balduzzi, Das, Foresi and Sundaram Model

In order to prove the accuracy of our bounds for a wider class of models, we consider a stochastic volatility (and long run mean) model. We use the following parameter values, proposed in Balduzzi et al. (1996) : $k = 0.25$, $\alpha = 0.76$, $\beta = 0.12$, $\eta = 0.02$, $a = 0.29$, $b = 0.0007$, $\phi = 0.003$ and $\rho = -0.12$.

Numerical results for this model are shown in Table 9.5.

9.5 Conclusions

This paper provides a new lower bound method for pricing of swaptions that is accurate, fast and applicable to a wide range of interest rate models. Our algorithm is particularly efficient because it requires the computation of only one Fourier inversion. Existing approximations (for instance Singleton and Umantsev method) require a number of Fourier inversions equal to the number of payment dates of the underlying swap. An approximate exercise region defined by the log-geometric mean of the ZCBs portfolio is tried out for the first time for swaption pricing. From the numerical tests we find that the approximation is much faster than Monte Carlo method and it is also very accurate across different maturities, tenors and strikes. Numerical results are presented across a wide class of models, including model with jumps and stochastic volatility. For all these models our lower bound is applicable and it is accurate, instead the Singleton and Umantsev method is not applicable if the density function is not known in analytical form. Moreover the lower bound is very effective as control variate to reduce the computation time and the error of the Monte Carlo. Hence, our model could be very suitable also for calibration purposes.

A Appendix

A.1 Proof Proposition 1

We consider the lower bound to the swaption price as in formula (9.4) for affine models:

$$LB(k; t, T, \{T_h\}_{h=1}^n, R) = P(t,T) \cdot \mathbb{E}_t^T \left[\left(\sum_{h=1}^n w_h \, e^{\mathbf{b}_h^\top \mathbf{X}(T) + a_h} - 1 \right) I(\mathcal{G}) \right]^+$$

where the set $\mathcal{G} = \{\omega \in \Omega : g(\mathbf{X}(T)) \geq k\} = \{\omega \in \Omega : \boldsymbol{\beta}^\top \mathbf{X}(T) + \alpha \geq k\}$.

According to Carr and Madan (2000), we introduce the dampening factor $e^{\delta k}$, then we apply the Fourier Transform with respect to the variable k to the T-forward expected value and we obtain:

$$\begin{aligned}
\psi_\delta(\gamma) &= \int_{-\infty}^{+\infty} e^{i\gamma k + \delta k} \mathbb{E}_t^T \left[\left(\sum_{h=1}^n w_h \, e^{\mathbf{b}_h^\top \mathbf{X}(T) + a_h} - 1 \right) I(g(\mathbf{X}(T)) \geq k) \right] dk \\
&= \mathbb{E}_t^T \left[\left(\sum_{h=1}^n w_h \, e^{\mathbf{b}_h^\top \mathbf{X}(T) + a_h} - 1 \right) \int_{-\infty}^{+\infty} e^{i\gamma k + \delta k} I(\boldsymbol{\beta}^\top \mathbf{X}(T) + \alpha \geq k) dk \right] \\
&= \mathbb{E}_t^T \left[\left(\sum_{h=1}^n w_h \, e^{\mathbf{b}_h^\top \mathbf{X}(T) + a_h} - 1 \right) \int_{-\infty}^{\boldsymbol{\beta}^\top \mathbf{X}(T) + \alpha} e^{i\gamma k + \delta k} dk \right]
\end{aligned}$$

Since the dampening factor δ is positive, then the module of the integrand function decays exponentially for $k \to -\infty$ and the Fourier Transform is well defined, so:

$$\psi_\delta(\gamma) = \mathbb{E}_t^T \left[\left(\sum_{h=1}^n w_h e^{\mathbf{b}_h^\top \mathbf{X}(T) + a_h} - 1 \right) e^{(i\gamma + \delta)(\boldsymbol{\beta}^\top \mathbf{X}(T))} \right] \frac{e^{(i\gamma+\delta)\alpha}}{i\gamma + \delta}$$

Using the characteristic function of $\mathbf{X}$, calculated under the T-forward measure: $\Phi(\boldsymbol{\lambda}) = \mathbb{E}_t^T \left[e^{i\boldsymbol{\lambda}^\top \mathbf{X}} \right]$, the function $\psi_\delta(\gamma)$ can be written as:

$$\psi_\delta(\gamma) = \left(\sum_{h=1}^n w_h e^{a_h} \Phi\left(-i\mathbf{b}_h + (\gamma - i\delta)\boldsymbol{\beta}\right) - \Phi\left((\gamma - i\delta)\boldsymbol{\beta}\right) \right) \frac{e^{(i\gamma+\delta)\alpha}}{i\gamma + \delta}$$

Finally the lower bound is the maximum with respect to k of the inverse transform of $\psi_\delta(\gamma)$:

$$LB(k; t, T, \{T_h\}_{h=1}^n, R) = P(t,T) \, \frac{e^{-\delta k}}{\pi} \int_0^{+\infty} e^{-i\gamma k} \psi_\delta(\gamma) d\gamma$$

A.2 Proof of the Analytical Lower Bound for Gaussian Affine Models

$\mathbf{X} \sim N(\boldsymbol{\mu}, V)$ in T-forward measure and $g(\mathbf{X}(T)) = \boldsymbol{\beta}^\top \mathbf{X} + \alpha \sim N(\boldsymbol{\beta}^\top \boldsymbol{\mu} + \alpha,\ \boldsymbol{\beta}^\top V \boldsymbol{\beta})$

Then the approximate exercise region $\mathcal{G}$ becomes:

$$\mathcal{G} = \{\omega \in \Omega : g(\mathbf{X}(T)) > k\} = \{\omega \in \Omega : z > d\}$$

where z is a standard normal random variable and

$$d = \frac{k - \boldsymbol{\beta}^\top \boldsymbol{\mu} - \alpha}{\sqrt{\boldsymbol{\beta}^\top V \boldsymbol{\beta}}}.$$

The lower bound expression can be written using the law of iterative expectation:

$$LB(k; t, T, \{T_h\}_{h=1}^n, R) = P(t,T)\, \mathbb{E}_t^T \left[\mathbb{E}_t^T \left[\left(\sum_{h=1}^n w_h e^{\mathbf{b}_h{}^\top \mathbf{X}(T) + a_h} - 1 \right) |z \right] I(z > d) \right]$$

Conditionally to the random variable z, the variable $\mathbf{X}$ is distributed as a multivariate normal with mean and variance:

$$\mathbb{E}_t^T[\mathbf{X}|z] = \boldsymbol{\mu} + z \cdot \mathbf{v}$$
$$Var(\mathbf{X}|z) = V - \mathbf{v}\mathbf{v}^\top$$
$$\mathbf{v} = \frac{V\boldsymbol{\beta}}{\sqrt{\boldsymbol{\beta}^\top V \boldsymbol{\beta}}}$$

We can now compute the inner expectation, using the normal distribution property:

$$LB(k; t, T, \{T_h\}_{h=1}^n, R) = P(t,T) \left(\sum_{h=1}^n w_h\, \mathbb{E}_t^T \left[e^{\mathbf{b}_h^\top \boldsymbol{\mu} + z \mathbf{b}_h^\top \mathbf{v} + \frac{1}{2} V_h} I(z > d) \right] - \mathbb{E}_t^T\left[I(z > d) \right] \right)$$

where $V_h = \mathbf{b}_h^\top (V - \mathbf{v}\mathbf{v}^\top)\mathbf{b}_h$.

Maximizing with respect to k, involved in the definition of d, we found the lower bound:

$$LB(t, T, \{T_h\}_{h=1}^n, R) = \max_{k \in \mathbb{R}} \left(\sum_{h=1}^n w_h e^{a_h + \mathbf{b}_h^\top \boldsymbol{\mu} + \frac{1}{2} V_h + \frac{1}{2} d_h^2} N(d_h - d) - N(-d) \right)$$

where $d_h = \mathbf{b}_h^\top \mathbf{v}$ and $N(x)$ is the cumulative distribution function of standard normal variable.

To make faster the optimization of the lower bound with respect to the parameter k, we compute the first order approximation of maximum point as a starting point. Equation for stationary points:

$$\frac{\partial LB(t,k)}{\partial k} = \frac{P(t,T)}{\sqrt{\boldsymbol{\beta}^\top V \boldsymbol{\beta}}} \left(\sum_{h=1}^{n} w_h e^{a_h + \mathbf{b}_h^\top \boldsymbol{\mu} + \frac{1}{2} V_h + d_h^2/2} N'(-(d-d_h)) - N'(-d) \right) = 0$$

Taylor first order expansion:

$$N'(-(d-d_h)) = N'(-d) + d_h N''(-d) + o(d_h) = N'(-d)(1 + dd_h) + o(d_h)$$

We substitute the first order expansion in the derivative expression and we obtain:

$$\sum_{h=1}^{n} w_h e^{a_h + \mathbf{b}_h^\top \boldsymbol{\mu} + \frac{1}{2} V_h + d_h^2/2} (1 + dd_h) - 1 = 0$$

So the first order guess of the maximum point is:

$$d_{guess} = \frac{1 - \sum_{h=1}^{n} w_h e^{a_h + \mathbf{b}_h^\top \boldsymbol{\mu} + \frac{1}{2} V_h + d_h^2/2}}{\sum_{h=1}^{n} w_h e^{a_h + \mathbf{b}_h^\top \boldsymbol{\mu} + \frac{1}{2} V_h + d_h^2/2} d_h}$$

$$k_{guess} = \sqrt{\boldsymbol{\beta}^\top V \boldsymbol{\beta}}\, d_{guess} + \boldsymbol{\beta}^\top \boldsymbol{\mu} + \alpha$$

References

P. Balduzzi, S.R. Das, S. Foresi, R. Sundaram, A simple approach to three-factor affine term structure models. J. Fixed Income **6**(3), 43–54 (1996)

R. Caldana, G. Fusai, A. Gnoatto, M. Grasselli, General closed-form basket option pricing bounds. Working paper, posted at ssrn.com in Jan 2014

P. Carr, D.P. Madan, Option valuation using the fast Fourier transform. J. Comput. Financ. **2**, 61–73 (2000)

P. Collin-Dufresne, R.S. Goldstein, Pricing swaptions within an affine framework. J. Deriv. **10**, 9–26 (2002)

D. Duffie, R. Kan, A yield-factor model of interest rates. Math. Financ. **6**, 379–406 (1996)

D. Duffie, P. Pan, K. Singleton, Transform analysis and asset pricing for affine jump-diffusions. Econometrica **68**, 1343–1376 (2000)

D. Duffie, K. Singleton, An econometric model of the term structure of interest-rate swap yields. J. Financ. **52**, 1287–1321 (1997)

C. Munk, Stochastic duration and fast coupon bond option pricing in multi-factor models. Rev. Deriv. Res. **3**, 157–181 (1999)

D.F. Schrager, A.J. Pelsser, Pricing swaptions and coupon bond options in affine term structure models. Math. Financ. **16**, 673–694 (2006)

K.J. Singleton, L. Umantsev, Pricing coupon-bond options and swaptions in affine term structure models. Math. Financ. **12**, 427–446 (2002)

Part III
Optimization Techniques

Chapter 10
Portfolio Optimization Using Modified Herfindahl Constraint

Asmerilda Hitaj and Giovanni Zambruno

Abstract Modern portfolio theory started with Markowitz (J Financ 7(1):77–91, 1952; Portfolio selection efficient diversification of investments. Wiley, New York, 1959). Early works developed necessary conditions on utility function that would result in mean-variance theory being optimal, see Tobin (Rev Econ Stud 25(2):65–86, 1958). Recently, considering the stylized facts of asset returns, mean-variance model has been extended to higher moments. Despite all, empirical evidence has shown that mean-variance model and its variants often yield overly concentrated portfolios. Portfolio diversification is still an open question. To avoid this problem different constraints have been introduced in the portfolio optimization procedure. In this paper we study from an empirical point of view the impact of imposing a constraint on the Modified Herfindahl index of the portfolio, in case of mean-variance and mean-variance-skewness optimization. We find that imposing a constraint on the level of the portfolio diversification leads to better out of sample performance and significant gains, despite the use of shrinkage estimators for moments and comoments, in particular when long estimation periods are considered.

Keywords Higher moments portfolio selection • *Risk-based* strategies • Expected utility • Herfindahl index • Diversity constraint

10.1 Introduction

Markowitz (1952, 1959) is unanimously recognized as the originator of modern portfolio theory, where the importance of portfolios, risk, correlations between securities and diversification is emphasized. Using Quadratic Programming, he

---

A. Hitaj (✉) • G. Zambruno
Dipartimento di Statistica e Metodi Quantitativi, University of Milano Bicocca,
Via Bicocca degli Arcimboldi 8, 20126, Milano, Italy
e-mail: asmerilda.hitaj1@unimib.it; giovanni.zambruno@unimib.it

G. Consigli et al. (eds.), *Handbook of Recent Advances in Commodity and Financial Modeling*, International Series in Operations Research & Management Science 257,
DOI 10.1007/978-3-319-61320-8_10

introduced a computation method for determining mean-variance efficient portfolios as portfolios with minimal risk for a given return, or, equivalently, with the highest return for a given level of risk. Merton in (1972) using Lagrange Multipliers obtained the efficient frontier analytically. Tobin in (1958) noted that mean-variance analysis is implied if probability distributions are Gaussian or if the user has a quadratic utility function. Alternative portfolio theories, based on expected utility theory and taking into account the stylized facts of asset returns, included higher moments in portfolio allocation: among others, see Jondeau et al. (2007), Hitaj and Mercuri (2013a), and Martellini and Ziemann (2010). Consequently, expected utility maximization became a predominant criterion for portfolio choice. In order to account for some of the stylized facts of asset returns, usually a Taylor expansion of the expected utility function is considered, stopped at the second, third or fourth order if one wishes to take into account, respectively, mean-variance, mean-variance-skewness or mean-variance-skewness-kurtosis. Therefore a crucial point, when dealing with portfolio choice, is the estimation of the arrays of expected returns, co-variances, co-skewnesses and co-kurtosis, depending on the model used. In practice their estimation is a difficult task: traditionally the historical (sample) estimation has been used, which might not be appropriate for future data. Many papers have shown that portfolio optimization is very sensitive to these inputs (see, for example, Frankfurter et al. (1971), Best and Grauer (1991a), and Chopra and Ziemba (1993), in the case of the mean-variance model). This sensitivity has generally been attributed to the tendency for optimization to magnify the effects of estimation error, see e.g. Michaud (1989). Moreover, empirical evidence has shown that the mean-variance portfolio is over-concentrated in a few assets, see Best and Grauer (1991b) and Green and Burton (1992), which goes against the idea of diversification. Therefore a common problem to solve for practitioners and academics is the portfolio robustness, namely the need to find solutions that are poorly sensitive to variations in input values. A vast literature is now available on how to deal with the problem. One solution is to use high frequency data, for example daily data in place of monthly data. However this solution is not always possible: e.g. in case of hedge fund portfolios daily observations are not available. Usually when hedge fund portfolios are considered we have to work with monthly observations. A different solution is the use of improved estimation procedures (see Jorion 1986; Ledoit and Wolf 2003) and model variations (see, among others, Jagannathan and Ma 2003; Jondeau et al. 2007; Martellini and Ziemann 2010). In Ledoit and Wolf (2003) only the covariance matrix is obtained using the shrinkage estimator. This problem has been extended to higher-moment portfolio allocation in Martellini and Ziemann (2010) and applied on actual market data in Hitaj et al. (2012).

In practice, institutional investors are often forced by regulations to diversify their portfolios. Therefore the idea of imposing diversification constraints is commonly used in the asset management industry. Some of them, as proposed in literature, are: upper/lower bounds on asset weights (see Peter et al. 1988); L^p-norm constraints (see Demiguel and Nogales 2009), entropy constraint as used in Huang (2012). Empirical studies show that, when sample estimators are used for the mean and covariance elements, weight constraints improve portfolio efficiency out-of-sample (see among others Jagannathan and Ma 2003; Demiguel and Nogales 2009; Peter et al. 1988). Notably, some diversification constraints are in the form of an upper bound, while others are instead in the form of a lower bound.

Jagannathan and Ma in (2003) proved that weight constraints possess a shrinkage-like effect, meaning that imposing weight constraints or using shrinkage estimators for moments and comoments lead to similar results. In a recent paper, Hitaj and Zambruno (2016), the authors have shown through empirical analysis that the level of risk aversion has an important role on portfolio diversification. For low levels of risk aversion the optimal portfolio is overly concentrated and as risk aversion increases portfolio diversification also increases; yet, in case of non-normal return distributions, despite the use of shrinkage estimators, the obtained portfolios are not very diversified.

The aim of this paper is to move one step further and analyze the effects that imposing a constraint on the level of portfolio diversification has in an out-of-sample perspective, when shrinkage estimators are considered for moments and comoments. To our knowledge the optimization problem that involves a constraint on the level of portfolio diversification, when shrinkage estimators are used for moments and comoments, has not been studied yet. Moreover, we perform a detailed analysis not only for the *mean-variance* model but also for *mean-variance-skewness* one. This latter model is recommended for use when assets returns are not normally distributed. For this reason, we consider three different portfolios of hedge fund indexes, taken from Credit Suisse Alternative indexes, Hedge Fund Research Alternative indexes, EDHEC Alternative indexes; for sake of comparison we consider also two equity portfolios. The overall time-series of returns spans January 1997 to January 2013, consisting in 193 monthly observations. In the empirical part, for portfolio allocation we use mean-variance (MV) and mean-variance-skewness (MVS) models, as we will see in Sect. 10.3.2, and for comparison purposes we estimate also the optimal portfolios obtained using some *risk-based* strategies (explained in Sect. 10.3.1) widely used in the equity world.

We analyze the impact that the estimation window length has on the obtained results, by considering different buy-and-hold rolling window strategies such as 24-month or 48-month in-sample and 3-month or 6-month out-of-sample. Furthermore we analyze the impact of the risk aversion parameter, by considering different values of this. Since it has long been recognized that mean-variance efficient portfolios established using the sample covariance matrix perform poorly out-of-sample (see among others Best and Grauer (1991a), Jorion (1986), and Ledoit and Wolf (2003) in case of MV allocation and Hitaj et al. (2012) in case of MVSK model), in this work we use *shrinkage toward the constant correlation* for covariances and coskewnesses and *shrinkage toward the grand mean* for the mean.

Our main empirical findings are: (i) Three-moment portfolio selection leads to better out-of-sample performances than simple mean-variance also when a constraint on the level of the portfolio diversification is imposed. (ii) key variables in the portfolio allocation procedure are not only the level of risk aversion but also the length of the estimation window. In general we observe that despite the use of shrinkage estimators for moments and comoments, imposing a constraint on the level of portfolio diversification leads to better out-of-sample performances with respect to only shortselling constraints. These results are more evident and become statistically significant when long estimation windows are considered.

The paper is organized as follows: Sect. 10.2 gives a brief description of the commonly used diversification constraints. Section 10.3 discusses the MV and the MVS portfolio selection models under weight constraints used in the empirical part. Section 10.4 describes the empirical analysis, where we investigate the characteristics of the optimal portfolios and discuss the results obtained; Sect. 10.5 draws some conclusions.

10.2 Review of the Constraints About Diversification

Consider an investor who selects his portfolio from n risky assets and denote with $\mathscr{M}$ the collection of all parameters used to describe the assets under analysis and let $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ be the vector of weights, where w_i is the fraction of the initial endowment invested in the i-th asset. In general, when short sales are not allowed, the investor problem has the following form:

$$\begin{cases} \max\limits_{\mathbf{w}} / \min f(\mathscr{M},\ \mathbf{w}) \\ \sum_{i=1}^{n} w_i = 1 \\ 0 \le w_i \end{cases}, \tag{10.1}$$

where the first constraint, $\sum_{i=1}^{n} w_i = 1$, requires that the whole wealth be invested; the second constraint, $0 \le w_i$, prevents short selling; $f(\mathscr{M},\ \mathbf{w})$ is the objective function that the investor should optimize. The well-known mean-variance model belongs to this class. It has long been recognized that the MV-efficient portfolios established using sample estimators perform poorly out-of-sample and are highly concentrated in a few assets. Considering that, by regulations, many institutional investors cannot concentrate their portfolios in only a few assets, different weight constraints have been proposed in order to obtain a diversified portfolio. We briefly describe here those most used in literature (see Jun-Lin 2013).

10.2.1 Upper-Bound Constraint

One way to avoid overly concentrated portfolios is by adding an upper-bound constraint U_w on weights, meaning that we cannot invest more than a fraction U_w of wealth in any asset.

$$\max_i (w_i) \le U_w. \tag{10.2}$$

If $U_w = \frac{1}{n}$ then the feasible region in (10.1) contains only one portfolio, namely the equally-weighted one which is the most diversified portfolio. If $U_w = 1$, the constraint (10.2) becomes redundant. Therefore, $U_w \in \left[\frac{1}{n}, 1\right]$. Decreasing the value U_w we shrink the feasible region in (10.1) forcing a higher level of diversification.

10.2.2 Lower-Bound Constraint

An alternative way to obtain a diversified portfolio is by imposing a lower bound (L_w) constraint on weights. We impose to invest no less than L_w in each asset:

$$\min_i (w_i) \geq L_w. \tag{10.3}$$

If $L_w = 0$, constraint (10.3) becomes redundant due to constraint in (10.1). If $L_w = \frac{1}{n}$ the feasible region in (10.1) contains only one portfolio, the equally-weighted one. Therefore $L_w \in \left[0, \frac{1}{n}\right]$. The higher L_w the more diversified is the portfolio.

Enforcing a weight lower-bound constraint automatically induces an upper-bound constraint: if $min(\mathbf{w}) \geq L_w$ the maximum weight invested in one asset is necessarily $1 - (n-1)L_w$.

10.2.3 L^p-Norm Constraint

For any $p > 1$, the L^p-norm constraint imposes an upper bound U_L on the L^p-norm ($L(\mathbf{w})$) of a portfolio $\mathbf{w}$, defined as:

$$L(\mathbf{w}) = \sum_{i=1}^{n} |w_i|^p \leq U_L. \tag{10.4}$$

For the least diversified portfolio, the one where all the wealth is invested in one asset, we obtain the maximum value: $L(\mathbf{w}) = 1$. For the most diversified portfolio, where $w_i = \frac{1}{n}$, *for* $i = 1, \ldots n$, $L(\mathbf{w})$ attains its lowest value $L(\mathbf{w}) = n^{1-p}$. Therefore $U_L \in \left[n^{1-p}, 1\right]$. Lower values of U_L shrink the feasible region in (10.1) to more diversified portfolios.

Lemma 2.1 *Imposing a L^p-norm constraint we implicitly impose an upper-bound constraint. In particular, if $L(\mathbf{w}) \leq U_L$ then $\max_i (w_i) \leq (U_L)^{\frac{1}{p}}$.*

Proof Proof by contradiction. Assume $\mathbf{w}$ has a component $w_i > (U_L)^{\frac{1}{p}}$. Then, being $p > 1$, we have $w_i^p > U_L$. Considering the no short sale constraint, $w_j \geq 0$ for all j, in (10.1), $\sum_{j \neq i} w_j \geq 0$. $L(\mathbf{w}) = w_i^p + \sum_{j \neq i} w_j{}^p > U_L$, which contradicts (10.4).

We remark that for $p = 2$ we obtain the Herfindahl index, which is often used as a measure of portfolio diversification. The Herfindahl index is defined as:

$$H_I = \sum_{i=1}^{n} w_i^2.$$

In the case of highest diversification, where all $w_i = \frac{1}{n}$, the Herfindahl index reaches its minimum value, $H_I = \frac{1}{n}$. In the case of highest concentration, where only one $w_i = 1$ and the rest are zero, the Herfindahl index reaches its maximum value 1. Therefore, $H_I \in \left[\frac{1}{n},\ 1\right]$. It is evident that the Herfindahl index, defined as such, depends on market size n.

10.2.4 Entropy Constraint

Also the *Shannon's entropy* ($Sh(\mathbf{w})$) of a portfolio $\mathbf{w}$ may be used, defined as:

$$SH(\mathbf{w}) = -\sum_{i=1}^{n} w_i \ \ln w_i.$$

The entropy constraint imposes a lower bound, L_{SH}, on the *Shannon's entropy* of the portfolio $\mathbf{w}$:

$$SH(\mathbf{w}) = -\sum_{i=1}^{n} w_i \ \ln w_i \geq L_{SH}. \tag{10.5}$$

$SH(\mathbf{w})$ reaches its maximum value when all $w_i = \frac{1}{n}$, in this case $SH(\mathbf{w}) = \ln(n)$. The infimum of $SH(\mathbf{w})$ is approached when $w_i \longrightarrow 1$ for one asset and $w_{j \neq i} \longrightarrow 0$ for all the others: in this case $SH(\mathbf{w}) \longrightarrow 0$. Therefore $L_{SH} \in (0, \ln(n)]$. The higher L_{SH} the more diversified is the portfolio.

As we can observe these constraints shrink the feasible region in (10.1) in different ways. On one extreme we obtain the equally-weighted portfolio (highest diversification) and on the other extreme we have a portfolio invested in one asset only (highest concentration). It would be interesting to investigate the relation among these constraints in case of portfolio allocation both from the theoretical and the empirical point of view. Such analysis falls beyond the scope of the present work: rather, here we focus on one particular weight constraint, a transformation of the L^p-norm, and we analyze empirically its effect on the out-of-sample performance.

10.3 Portfolio Allocation Models Under Consideration

Since the pioneering work of Markowitz (1952), based on the mean-variance framework, several portfolio selection models have been proposed. Especially in the equity framework different strategies have been suggested in contrast to capitalization-weighted. Some of these are: equally weighted[1] (EW), global minimum variance (GMV), equal risk contribution[2] (ERC) and maximum diversified[3] (MD). They have raised great interest both from market practitioners and academic researchers. These approaches are known as *risk-based* or *smart beta*[4] strategies and are assumed to be robust since they rely at most on the estimation of the covariance matrix ignoring the expected returns. Several empirical analysis (see among others DeMiguel et al. (2009); Maillard et al. (2010); Choueifaty and Coignard (2008)) have been performed in case of equity portfolios, concluding for the superiority of these strategies with respect to those based on mean and variance. Therefore, in the empirical part, we use these strategies as reference portfolios for the expected utility portfolio allocation based on two (MV) and three (MVS) moments. We consider different levels of risk aversion and two types of weight constraints; no short selling and a modified version of the Herfindahl index.

In the following we briefly explain the strategies used in the empirical part of this paper.

10.3.1 Risk-Based Strategies

The *risk-based* strategies considered in this work are:

10.3.1.1 Equally Weighted Portfolio (EW)

The equally weighted strategy consists in holding a portfolio with weight $1/n$ in each component. This strategy completely ignores the data and does not require any optimization or estimation procedure.

[1]DeMiguel et al. (2009) have shown through empirical analysis that the EW portfolio has never been outperformed systematically by the other 13 strategies considered in their paper.

[2]Maillard et al. (2010) compared the GMV, the EW and the ERC in an asset/sector allocation context and concluded that the ERC approach might be considered a good trade-off between the other two methods in terms of absolute risk level, risk budgeting and diversification.

[3]Choueifaty and Coignard (2008) proposed the maximum diversified ratio and empirically found that the MDP is more efficient ex post compared to the market capitalization-weighted benchmark, the MV and the EW.

[4]The term *smart beta* is popular to denote any strategy which attempts to take advantage of the benefits of traditional passive investment, but adds a source of outperformance in order to beat traditional market capitalization weighted indices. For more information on the *risk-based* strategies see e.g. Amenc et al. (2013) and the references therein.

10.3.1.2 The Shortsale-Constrained Global Minimum-Variance Portfolio (GMV)

The GMV portfolio is the one offering the minimum overall variance: therefore it doesn't take into account the expected returns of the assets. Mathematically the investor problem can be written as:

$$\begin{cases} \min\limits_{\mathbf{w}} \sigma_P^2 = \mathbf{w}' \, M_2 \, \mathbf{w} \\ \text{s.t. } \; \sum_{i=1}^n w_i = 1 \\ 0 \leq w_i \end{cases}, \tag{10.6}$$

where M_2 is the covariance matrix of the portfolio components.

10.3.1.3 The Shortsale-Constrained Equal Risk Contribution Portfolio (ERC)

Qian in (2006) introduces the equal risk contribution strategy, where weights are such that each asset provides the same contribution to portfolio risk.

Let us recall that the marginal risk contribution of asset i is defined as:

$$\partial_{w_i}\sigma_P = \frac{\partial \sigma_P}{\partial w_i} = \frac{(M_2 \, \mathbf{w})_i}{\sqrt{\mathbf{w}' \, M_2 \, \mathbf{w}}}.$$

Indicating with $\sigma_i(w) = w_i \partial_{w_i}\sigma_P$ the risk contribution of the ith asset, the portfolio risk can be seen as the sum of risk contributions: $\sigma_P = \sum_i^N \sigma_i(w)$ (see Maillard et al. 2010). Therefore a characteristic of the ERC strategy is that:

$$w_i \; \partial_{w_i}\sigma_P = w_j \; \partial_{w_j}\sigma_P \qquad \forall \; i, \; j$$

This property allows the design of an algorithm for the easy determination of the weights: actually we can minimize the sum of all squared deviations from the preceding equalities:

$$\begin{cases} \min\limits_{\mathbf{w}} \; \sum_{i=1}^n \; \sum_{j=1}^n \; \left(w_i (M_2 \, \mathbf{w})_i - w_j (M_2 \, \mathbf{w})_j \right)^2 \\ \text{s.t. } \; \sum_{i=1}^n w_i = 1 \\ 0 \leq w_i \end{cases}, \tag{10.7}$$

and we can reach a proper equal risk contribution position only if the minimum is zero.

10.3.1.4 The Shortsale-Constrained Maximum Diversified Portfolio (MDP)

The basic idea behind the maximum diversification approach is to construct a portfolio that maximizes the benefits from diversification. Choueifaty and Coignard proposed in (2008) the so-called diversification ratio (DR), which is the ratio of the weighted average volatility of the assets to the portfolio actual volatility, given by: $DR = \frac{\sum_{i=1}^{n} w_i \sigma_i}{\sqrt{\mathbf{w}' M_2 \mathbf{w}}}$. Since different asset classes are not perfectly correlated with each other, this ratio in general is greater than 1, by an amount which expresses the reduction of total risk attributable to diversification. The investor problem in the MDP approach is:

$$\begin{cases} \max\limits_{w_i} DR = \frac{\sum_{i=1}^{n} w_i \sigma_i}{\sqrt{\mathbf{w}' M_2 \mathbf{w}}} \\ \text{s.t.} \quad \sum_{i=1}^{n} w_i = 1 \\ 0 \leq w_i \end{cases} \tag{10.8}$$

10.3.2 *Taylor Approximation of the Expected Utility (EU) with Constraints on Portfolio Diversification*

Introducing higher moments in portfolio selection using the Taylor expansion of the expected utility function is widely used in Finance (see among others Jondeau et al. 2007; Hitaj and Mercuri 2013b; Hitaj et al. 2012 etc.). In the particular case that a negative exponential utility function is used, its Taylor expansion up to the third order is the following:

$$f(\mathbf{w}) = -e^{-\lambda(\mu \mathbf{w})} \left[1 + \frac{\lambda^2}{2} \mathbf{w}' M_2 \mathbf{w} - \frac{\lambda^3}{6} \mathbf{w}' M_3 (\mathbf{w} \otimes \mathbf{w}) \right] \tag{10.9}$$

where λ is the level of risk aversion. This function has to be maximized subject to the constraints as in (10.1). The inputs of this problem are the vector of means μ, the co-variance and the co-skewness matrixes, M_2 and M_3 respectively. In the case of the *risk-based* strategies (*ERC*, *GMV* and *MDP*) we only need to estimate the covariance matrix, M_2.

In the empirical part of this work, taking into account that sample estimators are characterized by high estimation error[5] and considering the results reported in different empirical analysis, see among others Jorion (1986) and Ledoit and Wolf (2003) and Martellini and Ziemann (2010), Hitaj et al. (2012), and Hitaj (2010)[6]

[5] The estimation error is high when the sample size is small. This is specifically the case of Hedge Funds which do not have a long history.

[6] These papers have demonstrated that using improved estimators lead to better out-of-sample performance compared to that obtained using sample estimators.

we decided to use shrinkage estimators.[7] In particular for the mean we use the shrinkage toward the grand mean (see Jorion 1986) and for the co-variance and the co-skewness we use the shrinkage toward the constant correlation approach, see e.g. Ledoit and Wolf (2003, 2004), Martellini and Ziemann (2010), and Hitaj et al. (2012).

10.3.2.1 Diversifying Portfolios Through Weight Constraint

In this section we briefly review portfolio allocation according to mean-variance-skewness under *shortsale-constraints* and *Modified Herfindahl index* constraint, that we analyze in the empirical part.

The Shortsale-Constrained for CARA Expected Utility

Under *shortsale-constraints* the investor has to solve the following optimization problem:

$$\begin{cases} \max\limits_{\mathbf{w}} - e^{-\lambda(\mu\mathbf{w})}\left[1 + \frac{\lambda^2}{2}\mathbf{w}'M_2\mathbf{w} - \frac{\lambda^3}{6}\mathbf{w}'M_3(\mathbf{w}\otimes\mathbf{w})\right] \\ \qquad\qquad \text{s.t.} \quad \sum_{i=1}^{n} w_i = 1 \\ \qquad\qquad\qquad 0 \le w_i \end{cases}, \tag{10.10}$$

The constraints are the same as in (10.1) where short selling is not allowed. This is a natural constraint since in practice short positions are difficult to implement, and in case of hedge fund portfolios they are not allowed at all.

As has been shown in Hitaj and Zambruno (2016), despite the use of shrinkage estimators for moments and comoments the obtained portfolios are still highly concentrated, in particular when asset returns are not normally distributed. This can still be a problem as by law many institutional investors cannot concentrate their portfolios[8] and usually managers do not like to invest most of their endowment in one asset. Therefore, taking into account the fact that many institutional investors are often restricted by law to diversify their portfolios, in this paper we move one step further and consider the portfolio selection problem when a constraint on the level of portfolio diversification is added.

[7]The shrinkage estimator is based on averaging two different models: a high-dimensional model with low bias and high variance (the sample estimator), and a lower-dimensional model with larger bias but smaller variance (the target estimator). The estimated co-moment using the shrinkage approach is given by $M_i^{shrinkage} = kF + (1-k)S$, where k is a coefficient called shrinkage intensity, S is the sample co-matrix and F is the target co-matrix. In this paper as target co-matrix (F) we use the constant correlation, proposed by Elton and Gruber in (1973).

[8]This limit is easier to understand if we consider a Credit portfolio and think in terms of *Credit Risk*.

Modified Herfindahl Index Constraint for CARA Expected Utility

For an easy interpretation of the results, in this paper we consider a modified version of the Herfindahl index called Modified Herfindahl index, which is defined as:

$$MH_I = \frac{\sum_{i=1}^{n} w_i^2 - \frac{1}{n}}{1 - \frac{1}{n}}. \tag{10.11}$$

In the case of the most diversified portfolio, $w_i = \frac{1}{n}$ for $i = 1, \ldots n$, the value of the Modified Herfindahl index is $MH_I = 0$, while in the case of highest concentration, $w_i = 1$ and $w_{j \neq i} = 0$ for $i,\ j = 1, \ldots n$, $MH_I = 1$. Therefore $MH_I \in [0,\ 1]$ and this index decreases as the portfolio degree of diversification increases. The reason for using the Modified Herfindahl instead of the Herfindahl index is that the first one is independent on the size of the portfolio n.

In order to obtain a diversified portfolio, in this paper, a limit on the Modified Herfindahl index of the portfolio $\mathbf{w}$ is imposed. Therefore, the problem that we solve in this case is:

$$\begin{cases} \max\limits_{\mathbf{w}} - e^{-\lambda(\mu\mathbf{w})} \left[1 + \frac{\lambda^2}{2}\mathbf{w}' M_2 \mathbf{w} - \frac{\lambda^3}{6}\mathbf{w}' M_3 (\mathbf{w} \otimes \mathbf{w})\right] \\ \qquad \text{s.t. } \sum_{i=1}^{n} w_i = 1 \\ \qquad MH_I \leq U_{MH} \\ \qquad w_i \geq 0 \end{cases}, \tag{10.12}$$

where $U_{MH} \in [0,\ 1]$. The smaller is the U_{MH} the more diversified the portfolio must be.

Lemma 3.1 *Imposing an upper-limit on the MH_I we implicitly impose a weight upper-bound constraint. In particular we have that if $MH_I \leq U_{MH}$ then $\max\limits_i (w_i) \leq \left[U_{MH}(1 - \frac{1}{n}) + \frac{1}{n}\right]^{\frac{1}{2}}$.*

Proof From definition (10.11) we have that $H_I = \left[MH_I(1 - \frac{1}{n}) + \frac{1}{n}\right]$ and from Lemma 2.1 we have that $\max\limits_i (w_i) \leq (U_{H_I})^{\frac{1}{2}}$. Therefore we have $\max\limits_i (w_i) \leq \left[U_{MH}(1 - \frac{1}{n}) + \frac{1}{n}\right]^{\frac{1}{2}}$.

As we have explained in Sect. 10.2 different constraints can be considered in order to force portfolio diversification. The reason for adding a constraint on the MH_I of the portfolio is that in this way one fixes exante the desired level of diversification.

In the empirical part we consider different levels of risk aversion $\lambda = 0.5 : 0.5 : 25$,[9] meaning that we have in total 50 values and $U_{MH} = 0.1 : 0.05 : 0.5$. We stress that the values of risk aversion λ that we consider are for purpose of illustration

[9] In what follows we use the notation (min: step: max).

only, since it is hard to determine the appropriate values of this parameter which represent a proper risk aversion.

Considering the fact that Problems (10.10) and (10.12) are non-convex, in the empirical part, in order to solve these problems we use the Optimization Algorithm *global search*[10] of MATLAB. In order to determine a good starting point for the *global search* algorithm we split each axis in 10 equal parts and then consider all the permutations among the n axis whose components sum to 1 for problem (10.10) and in addition satisfy the Modified Herfindahl constraint in case of problem (10.12). These points are candidates for the initial point of the *global search* algorithm. We evaluate the objective function in each of these points and select as starting point the one that gives the objective function the highest value. Therefore, the starting point is not (as is usual) the EW portfolio and is not the same for all portfolios.

10.4 Empirical Analysis

We now turn to discuss the empirical analysis. First we explain the characteristics of the datasets used and then the procedure considered in order to measure the magnitude of potential gains or losses that can be realized by an investor when adding a constraint on the portfolio diversification level instead of only short-sale constraints.

10.4.1 Description of the Data Base

In this empirical analysis we consider three different portfolios of hedge fund indexes, taken respectively from Credit Suisse Alternative indexes,[11] Hedge Fund Research Alternative indexes[12] and EDHEC Alternative indexes.[13] Each component in these three sets, represents a different strategy. For the sake of comparison, we also build two Equity portfolios. Their components are taken from $S\&P$ 500 index. In constructing the first portfolio we selected 14 equities whose return distributions are not normal: in particular equities with a high Jarque-Bera test value; we call this portfolio *Equity non-normal*. The second portfolio is also an equity portfolio but in this case we selected from the $S\&P$ 500 index the components that have low Jarque-Bera test value (in this case each equity distribution is close to the normal one) and

[10] For more information on the Global Search algorithm see *http://it.mathworks.com/help/gads/how-globalsearch-and-multistart-work.html*.

[11] For more information, see www.hedgeindex.com, where monthly data of the considered Alternative Indexes can be downloaded.

[12] For more information,www.hedgefundresearch.com.

[13] For more information, see www.edhec-risk.com.

we call this portfolio *Equity normal*. The components of each portfolio are listed in Table A.1. For all the assets the overall time-series of returns spans January 1997 to January 2013, consisting in 193 monthly[14] observations.

In Tables A.2, A.3, and A.4, we report some descriptive statistics for each time series of returns, on the whole period under consideration, for the Credit Suisse and the two equity portfolios. The statistics reported in these tables are 'annual mean', 'annual std', 'skewness', 'kurtosis', 'JB-test' and its 'p-value'.

We find, as was expected, that most Credit Suisse hedge fund indexes, with the notable exception of 'Managed Futures' are negatively skewed and display a positive excess kurtosis (see Table A.2). At 1% significance level, we reject the null hypothesis that the returns are normally distributed for all the hedge fund indexes except 'Managed Futures index'. Similar results are obtained for the other two hedge fund portfolios; owing to space constraints they will not be reported here but will be provided by the authors upon request.

As explained previously, the *Equity Normal* portfolio, see Table A.3, was selected in such a way that the distribution of returns is close to normal. The results reported in this table show that, at the 1% significance level, we do not reject the null hypothesis that the returns are normally distributed for all the components except 'SO UN' (which has a $JB - test = 16.855$). In case of the *Equity non-normal* portfolio, Table A.4, at the 1% significance level we reject the null hypothesis of normality for all the components of this portfolio except 'KR UN'.

10.4.2 Empirical Protocol

Our objective is to analyze the impact that adding a constraint on the portfolio diversification level has on the out-of-sample performances of the MV and MVS portfolios, when shrinkage estimators are used for moments and comoments.

We analyze the impact that the estimation window length has on the obtained results, by considering a buy-and-hold strategy[15] with different rolling window lengths: that is 24 or 48 months in-sample and 3 or 6 months out-of-sample. In total we have four different rolling window situations. Moreover in order to analyze the impact that risk aversion has on the results we also consider 50 different values for the risk aversion parameter $\lambda = 0.5 : 0.5 : 20$.

[14]The choice of monthly returns is due to the frequency of Hedge Funds data available in the databases from where we downloaded the returns.

[15]Buy and hold strategy means that we estimate the optimal weights in the in-sample period and keep these constant in the next out-of-sample period.

10.4.2.1 In-Sample

In an in-sample perspective we analyze the optimal portfolios in terms of diversification, using the Modified Herfindahl index (MH_I) in Eq. (10.11). This index is 0 for the EW portfolio (which is deemed to be the most diversified portfolio) and is 1 if the portfolio is concentrated in one asset.

10.4.2.2 Out-of-Sample

In order to assess the magnitude of potential gains that can be attained by an investor when using a constraint on the level of portfolio diversification in addition to no-short selling constraints, an out-of-sample analysis of their performances is implemented. To this purpose the Information Ratio (IR) (where the *risk-based* strategies are used as reference portfolios) and the Sharpe Ratio (Sh) are employed as Risk Adjusted Performance Measures (RAPM) in order to compare the different portfolios. For these measures we test whether the Information (Sharpe) Ratios of two strategies are statistically distinguishable by implementing the studentized bootstrap procedure, proposed in Ledoit and Wolf (2008), with a block size of 6.

- The Sharpe Ratio is defined as;

$$Sh = \frac{\overline{R}_P - R_f}{\sigma_P}$$

  This ratio measures the average return of a portfolio in excess of the risk-free rate (R_f), also called the risk premium, as a fraction of the portfolio total risk, measured by its standard deviation.
- Considering the fact that *risk-based* strategies are widely used in the equity world and different empirical analysis demonstrate that these strategies perform better than the MV model and its variants (see among others Maillard et al. 2010; Choueifaty and Coignard 2008; DeMiguel et al. 2009) we compare our portfolios with some of the *risk-based* strategies using the Information Ratio, defined as:

$$IR = \frac{\overline{R}_P - \overline{R}_{ref}}{\sigma(R_P - R_{ref})}.$$

  where $\overline{R}_{ref}$ is the average return of the reference portfolio. In the empirical part we use as reference portfolio the four *risk-based* strategies described previously. Once the reference portfolio is fixed, managers seek to maximize IR, i.e. to reconcile a high residual return and a low tracking error. This ratio allows to check that the risk taken by the manager in deviating from the reference portfolio is sufficiently rewarded.

Different RAPM are used in literature and in practice to rank portfolios. Once the RAPM is fixed, the best portfolio is the one that has the highest RAPM.

The drawback of *Sh* is its dependence from $\overline{R}_f$. As a proxy for the risk-free rate of interest (R_f) the literature suggests the use of *1 month* or *3 month* maturity U.S. Treasury Bills (see e.g. Deguest et al. 2013) or alternatively an exogenously given value (see e.g. Brennan (1998) who considered $R_f = 5\%$). In the empirical part of this paper, for illustration purposes, we set $R_f = 0\%$.

10.4.3 Results with MH_I-Constraint

When dealing with Problems (10.10) and (10.12), in this empirical analysis we consider different levels of $\lambda = 0.5 : 0.5 : 25$. For each portfolio under analysis and for each in-sample period we find first the optimal weights solving Problem (10.6), (10.7), (10.8), (10.10) (the equally weighted portfolio does not require any optimization). We keep the weights constant in the next out-of-sample period and at the end we compute the out-of-sample returns and roll-over the next window. Moving this way we calculate all the out-of-sample portfolio returns which are used to evaluate *Sh* and *IR*.

The case of short-sale constraints has been studied in Hitaj and Zambruno (2016) and the main results in that paper are: (i) Despite the use of shrinkage estimators, the *MV* and *MVS* portfolios are highly concentrated, in particular for the Hedge Fund portfolios; (ii) the higher the risk aversion parameter the more diversified is the portfolio. Nevertheless for all levels of risk aversion considered the *GMV* portfolio, which is the least diversified among the *risk-based* considered, is more diversified than every *MV* and *MVS* in all the datasets under consideration; (iii) the *MV* and *MVS* portfolios, independently from the risk aversion parameter, allow for gains, in terms of *IR* and *Sh*, with respect to the *risk-based* strategies. Moreover *MVS* leads to higher out-of-sample performance than *MV* and this difference becomes statistically significant when the estimation period is long.

In this paper we move one step further and analyze the impact that adding a constraint[16] on the level of portfolio diversification has on the out-of-sample performance. For space limitations only the results obtained in case of $U_{MH} = 0.1, 0.3, 0.5$ are discussed here. We indicate with MV_i and MVS_i, for $i = 1, 2, 3$ the results obtained solving Problem (10.12), where U_{MH} is set at, respectively, $0.1, 0.3, 0.5$ and with MV_0 and MVS_0 the results obtained in case of no-short selling, solving problem (10.10). As explained previously, the lower the value of U_{MH} the more diversified is the portfolio.

For empirical illustration we report the results obtained for all the portfolios in case of *Credit Suisse* and *Equity Normal* portfolios, respectively, in Figs. A.1 and A.2 for the rolling-window strategies $24 - 3$ and $48 - 3$.

In a first stage we compare the MV_0 portfolios obtained using short-selling constraints with those obtained adding a constraint on the level of portfolio diversification, indicated with MV_1, MV_2 and MV_3. Some results obtained in case

[16]We solve problem (10.12) considering different levels of $U_{MH} = 0.1 : 0.05 : 0.5$.

of the *Equity Non Normal* dataset are reported in Table A.6 for the rolling window $48-6$. Moreover we compare the MV_0 with the three moments portfolio selection MVS_i for $i = 1, 2, 3$. For illustrative purpose we report in Table A.7 results obtained for selected levels of λ for the *Equity Non Normal* dataset.

Commenting the results, we can first observe that for a fixed U_{MH} and a fixed level of risk aversion, independently from the rolling window strategy, the MVS_i portfolios in most cases display higher out-of-sample performances (*IR* or *Sh*) than the MV_i ones: this is in line with the existing literature claiming that introducing higher moments to portfolio selection leads to better out-of-sample performances, which seems to be true also under portfolio diversification constraint.

Moreover, comparing the portfolios obtained under short selling constraints with those obtained adding some constraint on its diversification level, we observe that the latter in general leads to higher out of-sample performances, in terms of *IR* and *Sh*, when long estimation periods are considered. For empirical illustration we report in Fig. A.1c, d the results of the Credit Suisse portfolio with rolling-window $48-3$. From these figures it is clear that, in terms of *IR* and *Sh*, adding some constraint on the portfolio diversification level (MV_i and MVS_i for $i = 1, 2, 3$) almost always leads to better results with respect to the portfolios obtained with only shortsale constraints (MV_0 and MVS_0) for a fixed level of risk aversion.

In case of short estimation periods, the results are not so clear for all the datasets considered. For instance we report, for illustrative purpose, in Fig. A.1a, b the results obtained for the Credit Suisse dataset with rolling-window is $24-3$. Here for only some levels of λ adding a constraint on the portfolio diversification level leads to better out-of-sample performance with respect to short-sale constraints alone.

The results of the *Equity Normal* portfolio are reported in Fig. A.2, where we observe that MV_0, MV_2 and MV_3 have close out-of-sample performances: this is not unexpected since the *Equity Normal* portfolio obtained in case of only short-selling constraints is a very diversified portfolio (see $(2,1)$ in Fig. A.2a, c where MH_I is reported). In fact, in case of MV_0 and MVS_0, the $MH_I \leq 0.5$ which means that imposing a diversification constraint with $U_{MH} = 0.5$ will not alter our results. In contrast, for $U_{MH} = 0.1$ the out-of-sample results of MV_1 and MVS_1 are different from MV_0 and MVS_0, and in case of long estimation window the out-of-sample performances of MV_1 and MVS_1 are almost always higher than those of MV_0 and MVS_0.

In order to understand if the gains (losses), in terms of *IR* or *Sh*, obtained imposing some constraint on the portfolio diversification level in addition to only short-selling constraints are statistically significant, we implement the studentized bootstrap procedure, proposed in Ledoit and Wolf (2008), with a block size of 6. For illustrative purposes we report in Tables A.6 and A.7 some of the results obtained in case of *Equity Non Normal* dataset and rolling window $48-6$. In Table A.6, we report the results obtained, for selected levels of risk aversion λ, where we test if the *IR* of strategy MV_0 is equal to that of the $MV_{i=1,2,3}$ one. In Table A.7 we report the results obtained while testing if the *IR* of MV_0 strategy is equal to that of $MVS_{i=1,2,3}$. The respective p-values are indicated in brackets and the asterisk indicates 10% significance.

In order to have a general idea of the results obtained, in Table A.5 we report, for each dataset and rolling-window, the relative frequencies at which the null hypothesis (that two strategies have the same IR) cannot be accepted at a confidence level of 10% over all the 50 levels of risk aversion considered. We observe that the probability of not accepting the null hypothesis increases as the in-sample and out-of-sample window lengths increase. This means that the out-of-sample performances (Information Ratio in this case) of MV_0 and $MV_{i=1,\,2,\,3}$ (MV_0 and $MVS_{i=1,\,2,\,3}$) are statistically distinguishable at a confidence level of 10%. Putting these results together with those obtained previously, we can say that adding some constraints on the level of portfolio diversification in general leads to better out-of-sample performances with respect to the portfolio obtained with only short-selling constraints, but these gains become statistically significant in case of long estimation window and when asset returns are not normally distributed.[17]

Similar results, which for space constraints are not reported, are obtained in case the Sharpe ratio is used to measure the out-of-sample performance.

Considering the fact that imposing weight constraints has a shrinkage-like effect as demonstrated in Jagannathan and Ma (2003), we conjecture that, when working with monthly data and asset returns that are not normally distributed, the sampling error associated with the estimated optimal shrinkage intensity proposed in Ledoit and Wolf (2003) and Martellini and Ziemann (2010) is rather large in particular for long estimation windows. To our opinion, this is due to the fact that by considering a buy-and-hold rolling window strategy we do not take into account possible scenario changes which are important for portfolio allocation. In case of a long estimation window it may happen that we have an economic regime change and considering older data in the estimation procedure may lead to estimation error of moments and comoments also in case of the shrinkage estimators. Therefore we can say that considering portfolio diversification constraints leads in general to financial gains in an out-of-sample perspective, and that these gains become statistically significant in case of long estimation windows. However these results depend on the value of the upper bound U_{MH} and the size of the estimation window: these parameters should be set in accordance with the specific empirical problem at hand.

10.5 Conclusions

The aim of this paper is to shed some light on the characteristics of hedge fund indexes portfolios and analyze the impact that imposing *Modified Herfindahl index* (MH_I) constraint has on the out-of-sample performance. For this purpose three different portfolios of hedge fund indexes are considered. For the sake of comparison, we also build two Equity portfolios, whose components are taken from *S&P* 500 index. The overall time-series of returns spans January 1997 to January 2013, consisting in 193 monthly observations. For the portfolio allocation we use

[17]In case of *Equity Normal* portfolio the differences obtained, in terms of IR and Sh, between MV_0 and $MV_{i=1,2,3}$ (MV_0 and $MVS_{i=1,2,3}$) are not statistically significant both for long and short estimation windows. These results are not reported for space limitation but can be provided by the authors upon request.

the expected utility theory, considering the Taylor expansion of the second (MV) and third (MVS) order. In order to reduce the estimation error inherent in the sample approach we use the shrinkage estimators for moments and comoments. In particular the shrinkage toward the grand mean is used for the estimation of the mean and the shrinkage toward the constant correlation is used for the covariance and coskewness elements.

We analyze the impact that risk aversion has on the results by considering 50 different values of the risk aversion coefficient λ. Furthermore we investigate the impact that the length of the estimation window has on the results by considering a buy-and-hold strategy with different lengths for the rolling window: that is 24 or 48 months in-sample and 3 or 6 months out-of-sample. In total we have 4 different rolling windows situations.

We find that, despite the use of improved estimators for moments and comoments, the optimal portfolios are concentrated on a few items, particularly when we work with hedge funds and with low levels of risk aversion. Equity portfolios are always more diversified than the hedge fund ones. Moreover, we find that, when shrinkage estimators are used, portfolio selection with higher-order moments (*MVS*) consistently dominates mean-variance from an out-of-sample perspective, in terms of Information and Sharpe Ratio, also under portfolio diversification constraints. Furthermore, we find that adding some constraint on the level of portfolio diversification leads to better and significant gains in terms of Information Ratio and Sharpe Ratio, with respect to the portfolio obtained in case of only *shortsale constraints*, when dealing with hedge fund portfolios and long in-sample period. However, how much we need to constrain MH_I in order to get better out-of-sample performances, is an empirical issue and depends on the dataset. These facts lead to suspect that the sampling error associated with the estimated optimal shrinkage intensities is still large, especially when working with long estimation periods and non normal asset returns. In these cases, imposing some constraint on the level of portfolio diversification leads to significant gains. The results obtained suggest that more work is needed in improving the estimation of moments and comoments and in understanding the effect that different parameter values have on the results obtained.

These results are new in the literature and possess important potential implications for hedge fund investors and fund of hedge funds managers. They show that not only an accurate analysis of the portfolio characteristics is important in order to choose an appropriate model for portfolio allocation, but also a careful analysis is needed in order to understand the impact that the model parameters have on the results. In particular, the results here reported show that despite the use of shrinkage estimators, proposed to overcome the limits of sample estimators for moments and comoments, the obtained portfolios are still concentrated in a few assets. Moreover adding some constraint on the level of portfolio diversification, in general leads to gains in an out-of-sample perspective with respect to the portfolios obtained with only shortselling constraints. These gains become statistically significant as the estimation window increases.

A Empirical Results

Table A.1 Components of each portfolio under analysis

Credit Suisse portfolio	HFRI: portfolio	EDHEC portfolio	SPX non normal	SPX normal
Hedge Fund Index’	FOF: Strategic Index’	‘Convertible Arbitrage’	‘FMC UN Equity’	‘AMD UN Equity’
Convertible Arbitrage Hedge Fund Index’	Fund of Funds Composite Index’	‘CTA Global’	‘XEL UN Equity’	‘APA UN Equity’
Dedicated Short Bias Hedge Fund Index’	Fund Weighted Composite Index’	‘Distressed Securities’	‘AIG UN Equity’	‘CMCSA UW Equity’
Emerging Markets Hedge Fund Index’	Fund Weighted Composite Index CHF’	‘Emerging Markets’	‘FITB UW Equity’	‘ED UN Equity’
Event Driven Hedge Fund Index’	Fund Weighted Composite Index GBP’	‘Equity Market Neutral’	‘KR UN Equity’	‘FDX UN Equity’
Event Driven Distressed Hedge Fund Index’	Fund Weighted Composite Index JPY’	‘Event Driven’	‘TEG UN Equity’	‘GIS UN Equity’
Event Driven Multi-strategy Hedge Fund Index’	Macro (Total) Index’	‘Fixed Income Arbitrage’	‘BEAM UN Equity’	‘JNJ UN Equity’
Event Driven Risk Arbitrage Hedge Fund Index’	Macro: Systematic Diversified Index’	‘Global Macro’	‘HBAN UW Equity’	‘L UN Equity’
Fixed Income Arbitrage Hedge Fund Index’	Relative Value (Total) Index’	‘Long/Short Equity’	‘THC UN Equity’	‘NUE UN Equity’
Global Macro Hedge Fund Index’	RV: Fixed Income-Convertible Arbitrage Index’	‘Merger Arbitrage’	‘LNC UN Equity’	‘PAYX UW Equity’
Long/Short Equity Hedge Fund Index’	RV: Fixed Income-Corporate Index’	‘Relative Value’	‘CI UN Equity’	‘PFE UN Equity’
Managed Futures Hedge Fund Index’	RV: Multi-strategy Index’	‘Short Selling’	‘NEM UN Equity’	‘SO UN Equity’
		‘Funds Of Funds’	‘F UN Equity’	‘LUV UN Equity’
			‘WMB UN Equity’	‘WAG UN Equity’

Table A.2 General statistics for Credit Suisse Hedge Fund portfolio. Almost all the components in this portfolio are characterized by negative skewness and kurtosis greater than 3. Checking the p-value and the JB-test we can say that the null hypothesis of normality, at a significance level of 1%, cannot be accepted for each component except the 'Managed Futures'

	Period under consideration *Jan*/1997 *to Jan*/2013					
	Hedge Fund Credit Suisse indexes					
Ticker	Annual mean	Annual STD	Skewness	Kurtosis	JB-test	P-value
Hedge Fund	0.072	0.071	−0.371	6.655	111.8776	<0.001
Convertible Arbitrage	0.067	0.074	−2.920	20.733	2802.9545	<0.001
Short Bias	−0.045	0.173	0.748	4.391	33.5682	<0.001
Emerging Markets	0.074	0.140	−1.338	10.051	457.3451	<0.001
Event Driven	0.077	0.065	−2.359	13.684	1096.9649	<0.001
Event Driven Distressed	0.083	0.066	−2.505	15.638	1486.1835	<0.001
Event Driven Multi-strategy	0.075	0.071	−1.859	10.429	555.0094	<0.001
Event Driven Risk Arbitrage	0.049	0.043	−1.044	7.553	201.8156	<0.001
Fixed Income Arbitrage	0.042	0.062	−4.803	37.802	10482.2108	<0.001
Global Macro	0.101	0.086	−0.208	8.544	248.5432	<0.001
Long/Short Equity	0.084	0.102	−0.080	6.266	85.9773	<0.001
Managed Futures	0.053	0.116	0.066	2.606	**1.3906**	**0.445**

Table A.3 General statistics for the *Equity Normal* portfolio, where its components have been selected in such a way that their skewness is close to 0 and kurtosis to 3, therefore the distribution of these equities approaches normality. This is confirmed by the results of the Jarque-Bera test and its p-value which indicate that the null hypothesis of normality, at a significance level of 1%, cannot be rejected for each asset except the 'SO UN'. This data has been downloaded from Bloomberg

	Period under consideration *Jan*/1997 *to Jan*/2013					
	Equity Normal portfolio					
Ticker	Annual mean	Annual STD	Skewness	Kurtosis	JB-test	P-Value
AMD UN	−0.099	0.693	−0.055	3.052	**0.119**	**0.5000**
APA UN	0.106	0.354	−0.073	3.438	**1.717**	**0.3680**
CMCSA UW	0.116	0.287	−0.240	3.051	**1.872**	**0.3375**
ED UN	0.041	0.172	−0.153	4.062	**9.825**	**0.0158**
FDX UN	0.094	0.295	0.124	3.904	**7.062**	**0.0319**
GIS UN	0.061	0.171	−0.281	3.381	**3.700**	**0.1173**
JNJ UN	0.068	0.192	−0.151	4.084	**10.177**	**0.0145**
L UN	0.063	0.269	−0.248	3.510	**4.065**	**0.0977**
NUE UN	0.080	0.361	−0.091	3.079	**0.319**	**0.5000**
PAYX UW	0.073	0.280	0.195	3.256	**1.756**	**0.3602**
PFE UN	0.042	0.237	−0.216	2.943	**1.523**	**0.4104**
SO UN	0.073	0.177	−0.022	4.447	16.855	0.0041
LUV UN	0.059	0.313	−0.183	3.239	**1.540**	**0.4065**
WAG UN	0.086	0.271	0.002	3.460	**1.703**	**0.3708**

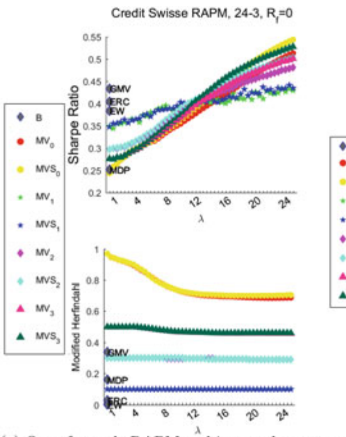
Table A.4 General statistics for *Equity non Normal* portfolio, where its component have been selected in such a way that their skewness is faraway from 0 and kurtosis faraway from 3, therefore the distribution of this equities is not Gaussian. This is confirmed by the results of the Jarque-Bera test and its p-value which indicate that the null hypothesis of normality, at a significance level of 1%, cannot be accepted for each equity in this portfolio except 'KR UN'

	Period under consideration *Jan*/1997 *to Jan*/2013					
	Equity non normal portfolio					
Ticker	Annual mean	Annual STD	Skewness	Kurtosis	JB-test	P-Value
FMC UN	0.117	0.323	−0.761	4.525	37.336	<0.001
XEL UN	0.012	0.302	−5.259	56.585	23980.108	<0.001
AIG UN	−0.151	0.774	−2.519	34.896	8385.386	<0.001
FITB UW	−0.008	0.493	−3.337	33.836	8004.620	<0.001
KR UN	0.054	0.240	−0.467	3.398	**8.303**	**0.023**
TEG UN	0.041	0.217	−3.488	33.079	7666.879	<0.001
BEAM UN	0.061	0.310	−0.357	9.604	354.854	<0.001
HBAN UW	−0.059	0.445	−2.775	24.278	3888.487	<0.001
THC UN	−0.025	0.606	−1.820	18.703	2089.577	<0.001
LNC UN	0.006	0.458	−1.557	17.057	1666.898	<0.001
CI UN	0.084	0.397	−2.464	17.032	1778.629	<0.001
NEM UN	−0.003	0.410	0.228	5.634	57.476	<0.001
F UN	−0.028	0.521	−0.416	13.417	878.205	<0.001
WMB UN	0.057	0.501	−1.696	12.285	785.845	<0.001

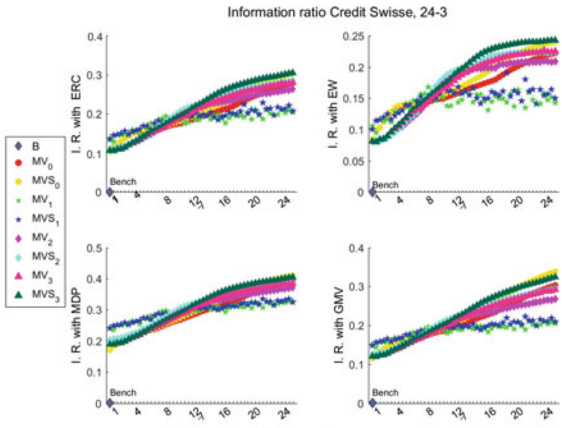
Table A.5 This table contains the relative frequencies at which $H_0 : IR_{MV_0} - IR_{MV_i} = 0$ ($H_0 : IR_{MV_0} - IR_{MVS_i} = 0$) cannot be accepted at a confidence level of 10%, over all levels (50) of risk aversion λ considered in our analysis. For example when the *Credit Suisse* dataset is considered, with a rolling window (RW) $48-6$ and *ERC* as reference portfolio, in 58% of the values of λ are such that the difference between IR_{MV_0} and IR_{MVS_3} is statistically significant at a confidence level of 10%. As we can observe, the relative frequency of not accepting H_0 almost always increases as the length of the estimation window increases

	Credit Suisse								EDHEC							
	Information ratio 24-3				Information ratio 24-6				Information ratio 24-3				Information ratio 24-6			
Benchmarks:	ERC	EW	MDP	GMV	ERC	EW	MDP	GMV	ERC	EW	MDP	GMV	ERC	EW	MDP	GMV
$MV_0 - MV_1$	0%	0%	0%	0%	0%	0%	0%	0%	8%	48%	2%	4%	0%	16%	0%	0%
$MV_0 - MV_2$	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%
$MV_0 - MV_3$	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%
$MV_0 - MVS_1$	0%	0%	0%	0%	0%	0%	0%	0%	4%	52%	0%	2%	0%	0%	0%	0%
$MV_0 - MVS_2$	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
$MV_0 - MVS_3$	0%	0%	0%	0%	38%	36%	38%	38%	0%	0%	0%	0%	16%	12%	14%	18%
	Information ratio 48-3				Information ratio 48-6				Information ratio 48-3				Information ratio 48-6			
Benchmarks:	ERC	EW	MDP	GMV	ERC	EW	MDP	GMV	ERC	EW	MDP	GMV	ERC	EW	MDP	GMV
$MV_0 - MV_1$	6%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	4%
$MV_0 - MV_2$	28%	0%	20%	2%	36%	22%	28%	30%	0%	0%	0%	0%	0%	0%	0%	0%
$MV_0 - MV_3$	44%	46%	48%	46%	48%	50%	46%	48%	0%	6%	0%	0%	20%	10%	0%	30%
$MV_0 - MVS_1$	16%	0%	0%	0%	6%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%
$MV_0 - MVS_2$	46%	34%	38%	42%	44%	34%	38%	44%	0%	0%	0%	0%	0%	0%	0%	0%
$MV_0 - MVS_3$	50%	52%	54%	56%	58%	56%	56%	54%	0%	24%	0%	0%	34%	34%	34%	36%

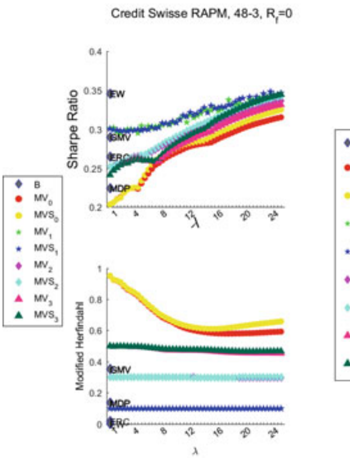
$MV_0 - MV_1$	2%	0%	2%	0%	0%	0%	0%	0%	78%	22%	82%	46%	56%	0%	80%	20%
$MV_0 - MV_2$	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	36%	0%	0%	0%	18%
$MV_0 - MV_3$	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
$MV_0 - MVS_1$	0%	0%	2%	0%	0%	0%	2%	0%	66%	22%	64%	38%	58%	8%	90%	30%
$MV_0 - MVS_2$	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	44%	0%	0%	0%	28%
$MV_0 - MVS_3$	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	4%	0%	0%	0%	0%
	Information ratio 48-3				Information ratio 48-6				Information ratio 48-3				Information ratio 48-6			
Benchmarks:	ERC	EW	MDP	GMV	ERC	EW	MDP	GMV	ERC	EW	MDP	GMV	ERC	EW	MDP	GMV
$MV_0 - MV_1$	8%	0%	0%	0%	46%	12%	0%	0%	16%	6%	24%	2%	32%	6%	38%	4%
$MV_0 - MV_2$	0%	0%	0%	0%	0%	0%	0%	0%	8%	0%	16%	2%	0%	0%	0%	0%
$MV_0 - MV_3$	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
$MV_0 - MVS_1$	6%	0%	0%	0%	42%	12%	0%	0%	22%	14%	26%	2%	20%	6%	28%	8%
$MV_0 - MVS_2$	0%	0%	0%	0%	2%	0%	0%	0%	62%	32%	64%	6%	42%	20%	40%	16%
$MV_0 - MVS_3$	0%	0%	0%	0%	2%	20%	18%	0%	0%	0%	0%	0%	66%	34%	66%	56%



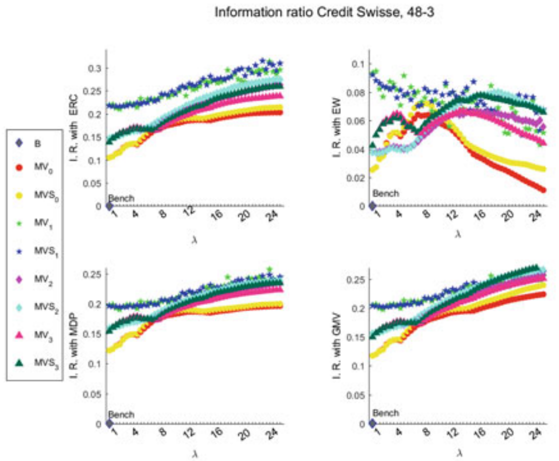
(a) Out-of-sample RAPM and in-sample average Modified Herfindahl.



(b) Out-of-sample IR.

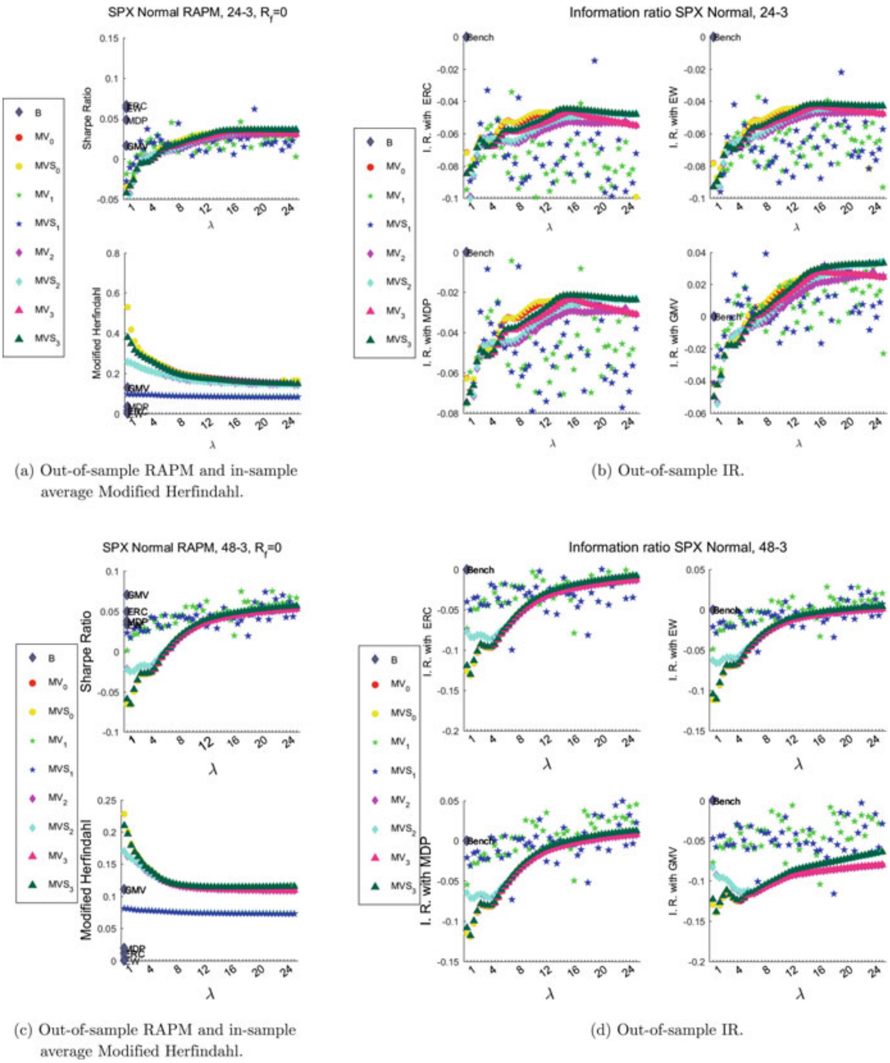


(c) Out-of-sample RAPM and in-sample average Modified Herfindahl.



(d) Out-of-sample IR.

Fig. A.1 Credit Suisse portfolio using expected utility approach with a rolling window of $24-3$ (in **a** and **b**) and $48-3$ (in **c** and **d**). Subfigures (**a**) and (**c**) report RAPM of MV_i and MVS_i; Sharpe in $(1,1)$ and Modified Herfindahl in $(2,1)$. MV_0 and MVS_0 are the results obtained solving Problem (10.10), while (MV_1 MVS_1), (MV_2 MVS_2) and (MV_3 MVS_3) are the results obtained solving Problem (10.12), respectively, for $U_{MH} = 0.1,\ 0.3$ and 0.5. Subfigures (**b**) and (**d**) report IR of MV_i and MVS_i portfolios with respect to ERC in $(1,1)$, EW in $(1,2)$, MDP in $(2,1)$ and GMV in $(2,2)$. As we can observe from subfigures $(2,1)$ in (**a**) and (**c**) the MV_0 and MVS_0 portfolios are highly concentrated portfolios, in particular for low levels of risk aversion. The MH_I for these portfolios ranges between 1 and 0.6



(a) Out-of-sample RAPM and in-sample average Modified Herfindahl.

(b) Out-of-sample IR.

(c) Out-of-sample RAPM and in-sample average Modified Herfindahl.

(d) Out-of-sample IR.

Fig. A.2 *S & P Normal portfolio* using Expected Utility approach with a rolling window strategy of $24-3$ (**a** and **b**) and $48-3$ (in **c** and **d**). Subfigures (**a**) and (**c**) report RAPM of MV_i and MVS_i; Sharpe in $(1, 1)$ and Modified Herfindahl in $(1, 2)$. MV_0 and MVS_0 are the results obtained solving Problem (10.10), while (MV_1 MVS_1), (MV_2 MVS_2) and (MV_3 MVS_3) are the results obtained solving Problem (10.12), respectively, for $U_{MH} = 0.1,\ 0.3$ and 0.5. Subfigures (**b**) and (**d**) report IR of MV_i and MVS_i portfolios with respect to ERC in $(1, 1)$, EW in $(1, 2)$, MDP in $(2, 1)$ and GMV in $(2, 2)$.As we can observe from subfigures $(2, 1)$ in (**a**) and (**c**) the MV_0 and MVS_0 portfolios are not concentrated portfolios. The MH_I for these portfolios ranges between; 0.6 and 0.2 for the rolling window $24-3$ (Subfigures $(2, 1)$ in **a**), 0.25 and 0.1 for the rolling window $48-3$ (Subfigures $(2, 1)$ in **c**)

Table A.6 Comparison between MV_0 and MV_i: results for the rolling window $48 - 6$ in case of *Equity Non Normal* dataset. This table reports the monthly out-of-sample Information Ratio of the *MV*, with respect to each Smart Beta as reference portfolio (indicated in bold), in case of no-short selling (labeled with MV_0) and in case of adding some constraints on the level of portfolio diversification (labeled with MV_i for $i = 1, 2, 3$, when the U_{MH} is respectively, 0.1, 0.3 and 0.5). In an out-of-sample perspective, for different levels of λ we compare the MV_0 with the MV_i (for $i = 1, 2, 3$) and we also test for the null hypothesis that the Information Ratio of the two strategies (MV_0 and MV_i) is equal. The respective p-value, indicated in brackets, is computed using the studentized circular bootstrap proposed in Ledoit and Wolf (2008) with a block size of 6. Asterisks denote 10% significance. For instance in case of $\lambda = 5$ and *ERC* as reference portfolio, the Information Ratio of the MV_0 strategy is 0.049 while that of the MV_1 is 0.117. Testing for the null assumption that the difference between these two Information Ratios is zero we obtain a p-value of 0.106 (at a level of 10% we do not reject the null hypothesis)

	Equity Non Normal Information Ratio; comparing MV_0 with MV_i							
	ERC				EW			
Benchmark:	MV_0	MV_1	MV_2	MV_3	MV_0	MV_1	MV_2	MV_3
$\lambda = 1$	0.089	0.098	0.066	0.083	0.050	0.034	0.025	0.045
		(0.894)	(0.369)	(0.611)		(0.704)	(0.341)	(0.674)
$\lambda = 5$	0.049	0.117	0.052	0.047	0.013	0.047	0.012	0.011
		(0.106)	(0.847)	(0.349)		(0.343)	(0.958)	(0.408)
$\lambda = 10$	0.042	0.098	0.048	0.042	0.004	0.026	0.008	0.004
		(0.129)	(0.460)	(0.85)3		(0.431)	(0.572)	(0.947)
$\lambda = 15$	0.039	0.112	0.046	0.039	0.001	0.039	0.005	0.001
		(0.033)*	(0.412)	(0.901)		(0.126)	(0.490)	(0.963)
$\lambda = 20$	0.039	0.119	0.047	0.039	0.000	0.039	0.005	0.000
		(0.331)	(0.289)	(0.900)		(0.422)	(0.339)	(0.956)
$\lambda = 25$	0.036	0.078	0.043	0.036	−0.003	0.013	0.002	−0.003
		(0.200)	(0.342)	(0.905)		(0.459)	(0.392)	(0.958)
Benchmark:	MDP				GMV			
$\lambda = 1$	0.096	0.108	0.074	0.091	0.117	0.137	0.101	0.111
		(0.851)	(0.401)	(0.784)		(0.492)	(0.534)	(0.658)
$\lambda = 5$	0.057	0.127	0.060	0.054	0.086	0.157	0.096	0.084
		(0.09)*	(0.843)	(0.377)		(0.110)	(0.446)	(0.348)
$\lambda = 10$	0.050	0.113	0.056	0.050	0.095	0.143	0.102	0.095
		(0.087)*	(0.440)	(0.850)		(0.388)	(0.559)	(0.944)
$\lambda = 15$	0.048	0.126	0.055	0.048	0.112	0.168	0.119	0.112
		(0.022)*	(0.402)	(0.896)		(0.294)	(0.623)	(0.954)
$\lambda = 20$	0.047	0.136	0.056	0.048	0.124	0.122	0.134	0.125
		(0.307)	(0.286)	(0.897)		(0.969)	(0.447)	(0.962)
$\lambda = 25$	0.045	0.093	0.052	0.045	0.130	0.139	0.138	0.130
		(0.125)	(0.334)	(0.899)		(0.867)	(0.554)	(0.959)

Table A.7 Comparison between MV_0 and MVS_i: Results for the rolling window 48-6 in case of *Equity Non Normal* dataset. This table reports the monthly out-of-sample Information Ratio of the MV_0 (in case of no-short selling) and MVS_i (in case of imposing a constraints on portfolio diversification level. MVS_i for $i = 1, 2, 3$, when the U_{MH} is respectively, 0.1, 0.3 and 0.5), with respect to each Smart Beta as reference portfolio (indicated in bold). In an out-of-sample perspective, for different levels of λ we compare the MV_0 with the MVS_i (for $i = 1, 2, 3$) and we also test for the null hypothesis that the Information Ratio of the two strategies (MV_0 and MVS_i) is equal. The respective p-value is indicated in brackets. Asterisks denote 10% significance. For instance in case of $\lambda = 15$ and EW as reference portfolio, we have that the Information Ratio of the MV_0 strategy is 0.001 while that of the MVS_2 is 0.014. Testing for the null assumption that the difference between these two Information Ratios is zero we obtain a p-value of (0.099)*. Therefore at a confidence level of 10% we reject the null hypothesis that the *IR* of the two strategies (MV_0 and MVS_2) are the same

	Equity Non Normal Information Ratio; comparing MV_0 with MVS_i							
	ERC				EW			
Benchmark:	MVS_0	MVS_1	MVS_2	MVS_3	MVS_0	MVS_1	MVS_2	MVS_3
$\lambda = 1$	0.089	0.098	0.066	0.083	0.050	0.034	0.025	0.045
		(0.854)	(0.377)	(0.597)		(0.670)	(0.340)	(0.666)
$\lambda = 5$	0.049	0.172	0.054	0.050	0.013	0.035	0.014	0.013
		(0.224)	(0.729)	(0.875)		(0.575)	(0.936)	(0.852)
$\lambda = 10$	0.042	0.113	0.053	0.048	0.004	0.036	0.012	0.009
		(0.075)*	(0.160)	(0.082)*		(0.244)	(0.240)	(0.107)
$\lambda = 15$	0.039	0.104	0.059	0.048	0.001	0.032	0.014	0.007
		(0.044)*	(0.061)*	(0.083)*		(0.172)	(0.099)*	(0.112)
$\lambda = 20$	0.039	0.089	0.059	0.046	0.000	0.021	0.014	0.006
		(0.422)	(0.078)*	(0.198)		(0.553)	(0.103)	(0.216)
$\lambda = 25$	0.036	0.094	0.056	0.048	−0.003	0.028	0.011	0.006
		(0.052)*	(0.074)*	(0.051)*		(0.141)	(0.108)	(0.064)*
Benchmark:	MDP				GMV			
$\lambda = 1$	0.096	0.105	0.074	0.090	0.117	0.137	0.102	0.111
		(0.834)	(0.382)	(0.797)		(0.495)	(0.536)	(0.648)
$\lambda = 5$	0.057	0.125	0.063	0.057	0.086	0.154	0.099	0.087
		(0.185)	(0.755)	(0.884)		(0.247)	(0.307)	(0.884)
$\lambda = 10$	0.050	0.128	0.062	0.057	0.095	0.163	0.109	0.103
		(0.058)*	(0.162)	(0.098)*		(0.248)	(0.140)	(0.082)*
$\lambda = 15$	0.048	0.118	0.068	0.057	0.112	0.158	0.137	0.124
		(0.035)*	(0.069)*	(0.090)*		(0.401)	(0.112)	(0.081)*
$\lambda = 20$	0.047	0.104	0.069	0.055	0.124	0.127	0.152	0.136
		(0.383)	(0.074)*	(0.187)		(0.981)	(0.288)	(0.172)
$\lambda = 25$	0.045	0.106	0.065	0.057	0.130	0.167	0.159	0.150
		(0.043)*	(0.067)*	(0.053)*		(0.458)	(0.254)	(0.472)

References

N. Amenc, F. Goltz, L. Martellini, Smart beta 2.0, June 2013

M.J. Best, R.R. Grauer, On the sensitivity of mean–variance efficient portfolios to changes in asset means: some analytical and computational results. Rev. Financ. Stud. **4**, 315–342 (1991)

M.J. Best, R.R. Grauer, Sensitivity analysis for mean-variance portfolio problems. Manag. Sci. **37**(8), 980–989 (1991)

M.J. Brennan, The role of learning in dynamic portfolio decisions. Eur. Financ. Rev. **1**, 295–306 (1998)

V.K. Chopra, W.T. Ziemba, The effect of errors in means, variances, and covariances on optimal portfolio choice. J. Portf. Manag. **19**(Winter), 6–11 (1993)

Y. Choueifaty, Y. Coignard, Towards maximum diversification. J. Portf. Manag. **35**(1), 40–51 (2008)

R. Deguest, L. Martellini, A. Meucci, Risk parity and beyond – from asset allocation to risk allocation decisions. Working paper (2013)

V. Demiguel, F.J. Nogales, A generalized approach to portfolio optimization: improving performance by constraining portfolio norms. Manag. Sci. **55**, 798–812 (2009)

V. DeMiguel, L. Garlappi, R. Uppal. Optimal versus naive diversification: how inefficient is the 1-n portfolio strategy? Rev. Financ. Stud. **22**(5), 1915–1953 (2009)

E.J. Elton, M.J. Gruber, Estimating the dependence structure of share prices-implications for portfolio selection. J. Financ. **28**(5), 1203–1232 (1973)

G.M. Frankfurter, H.E. Phillips, J.P. Seagle, Portfolio selection: the effects of uncertain means, variances, and covariances. J. Financ. Quant. Anal. **6**, 1251–1262 (1971)

R. Green, H. Burton, When will mean-variance efficient portfolios be well diversified? J. Financ. **47**(5), 1785–809 (1992)

A. Hitaj, Portfolio allocation under general return distribution. PhD thesis, University of Milano-Bicocca (2010)

A. Hitaj, L. Mercuri, Hedge fund portfolio allocation with higher moments and MVG models, in *Advances in Financial Risk Management: Corporates, Intermediaries and Portfolios*, ed. by J.A. Batten, P. Peter MacKay, N. Wagner (Palgrave Macmillan, New York, 2013a), pp. 331–346

A. Hitaj, L. Mercuri, Portfolio allocation using multivariate variance gamma models. Financ. Mark. Portf. Manag. **27**(1), 65–99 (2013b)

A. Hitaj, G. Zambruno, Are smart beta strategies suitable for hedge funds portfolios? Rev. Financ. Econ. **29**, 37–51 (2016)

A. Hitaj, L. Martellini, G. Zambruno, Optimal hedge fund allocation with improved estimates for coskewness and cokurtosis parameters. J. Altern. Invest. **14**(3), 6–16 (2012)

X.X. Huang, An entropy method for diversified fuzzy portfolio selection. Int. J. Fuzzy Syst. **14**, 160–165 (2012)

R. Jagannathan, T. Ma, Risk reduction in large portfolios: why imposing the wrong constraints helps. J. Financ. **58**(4), 1651–1684 (2003)

E. Jondeau, S. Poon, Rockinger M, *Financial Modeling Under Non-Gaussian Distributions. Springer Finance* (Springer, London, 2007)

P. Jorion, Bayes-Stein estimation for portfolio analysis. J. Financ. Quant. Anal. **21**(3), 279–292 (1986)

O. Ledoit, M. Wolf, Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. J. Empir. Financ. **10**(5), 603–621 (2003)

O. Ledoit, M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices. J. Multivar. Anal. **88**(2), 365–411 (2004)

O. Ledoit, M. Wolf, Robust performance hypothesis testing with the Sharpe ratio. J. Empir. Financ. **15**(5), 850–859 (2008)

J.-L. Lin, On the diversity constraints for portfolio optimization. Entropy **15**(11), 4607 (2013)

S. Maillard, T. Roncalli, J. Teiletche, On the properties of equally-weighted risk contributions portfolios. J. Portf. Manag. **36**(4), 60–70 (2010)

H. Markowitz, Portfolio selection. J. Financ. **7**(1), 77–91 (1952)
H. Markowitz, *Portfolio Selection Efficient Diversification of Investments* (Wiley, New York, 1959)
L. Martellini, V. Ziemann, Improved estimates of higher-order comoments and implications for portfolio selection. Rev. Financ. Stud. **23**(4), 1467–1502 (2010)
R.C. Merton, An analytic derivation of the efficient portfolio frontier. J. Financ. Quant. Anal. **7**(4), 1851–1872 (1972)
R.O. Michaud, The Markowitz optimization enigma: is optimized optimal. Financ. Anal. J. **1989**, 43–54 (1989)
P.A. Peter, A. Frost, J.E. Savarino, For better performance: constrain portfolio weights. J. Portf. Manag. **15**, 29–34 (1988)
E. Qian, On the financial interpretation of risk contribution: risk budgets do add up. J. Invest. Manag. **4**(4), 41–51 (2006)
J. Tobin, Liquidity preference as behavior toward risk. Rev. Econ. Stud. **25**(2), 65–86 (1958)

Chapter 11
Dynamic Asset Allocation with Default and Systemic Risks

Alessandro Sbuelz

Abstract Systemic risk breeds default risk. I investigate the optimal portfolio implications of their joint presence for non-myopic investors in arbitrage-free markets when such risks take the form of asset value discontinuities. I contribute to the multiple-asset jump-diffusion portfolio analysis of Das and Uppal (J Financ 59:2809–2834, 2004) by introducing default risk and its investment-horizon effects on optimal portfolios (the optimal investment rules in Das and Uppal (J Financ 59:2809–2834, 2004) are time-invariant) and by linking excess expected returns to risk exposures.

Keywords Strategic asset allocation • Investment-horizon effects • Investment opportunity set • Default risk • Systemic risk • Arbitrage-free markets • Risk premia • Jump-diffusive processes

JEL: G01, G11, G12, C61

11.1 Introduction

Default risk is closely related to systemic risk. Examples of how the default of a number of investible securities is associated with the simultaneous downward adjustment in the values of many tradable assets include but are not limited to the notable events of Fall 2008. The joint impact of default risk and systemic risk deeply affect investors. Importantly, as non-myopic rational investors anticipate the effect on their indirect utilities of the implosion of the defaultable securities

---

The usual disclaimer applies.

A. Sbuelz (✉)
Department of Mathematical Sciences, Mathematical Finance and Econometrics, Catholic University of Milan, Largo Gemelli 1, 20123, Milan, Italy
e-mail: alessandro.sbuelz@unicatt.it

G. Consigli et al. (eds.), *Handbook of Recent Advances in Commodity and Financial Modeling*, International Series in Operations Research & Management Science 257,
DOI 10.1007/978-3-319-61320-8_11

in a future systemic event, an investment-horizon effect arises even if the pre-default investment opportunity set is constant. As importantly, no-arbitrage markets command an additional excess-expected-return compensation for those twin risks that modifies the investment opportunity set itself.

I explore the dynamic portfolio implications of the joint presence of default risk and systemic risk in no-arbitrage multiple-asset markets with a constant opportunity set. Downward jump-like value adjustments are asset-specific but systemic (happen simultaneously). I extend the multiple-asset jump-diffusion portfolio analysis[1] of Das and Uppal (2004) in two important aspects. First, I focus on the default risk associated with systemic events and its investment-horizon effects on optimal portfolios (the optimal investment rules in Das and Uppal (2004) are time-invariant). Second, my model captures the dependence of the excess expected returns on the systemic jump sizes implied by the absence of arbitrage opportunities (Das and Uppal (2004) do not impose no-arbitrage constraints on the asset values dynamics).

This paper is related to the applied literature[2] that deals with contagion effects in continuous-time portfolio problems. Kraft and Steffensen (2009) study the asset allocation problem with several defaultable corporate bonds.[3] They do not focus on systemic risk as they rule out the possibility of a simultaneous downward jump-like correction in asset values. Branger et al. (2009, 2014) consider systemic risk with full/partial information but rule out default risk, whereas Puopolo (2015) focuses on default risk and transaction costs in the absence of systemic risk. This is also the case for the portfolio choice literature that characterizes contagion by means of regime-switching models (e.g. Ang and Bekaert 2002; Kole et al. 2006; Guidolin and Timmermann 2007, 2008) and by means of correlation risk without price discontinuities (e.g. Buraschi et al. 2010). Pagliarani and Vargiolu (2014) restrict the dynamic-asset-allocation analysis of default risk and systemic risk to the log-utility investor.

Section 11.2 outlines the arbitrage-free dynamics of the risky asset values. Section 11.3 discusses the optimal investment rule and Sect. 11.4 provides a numerical analysis. Section 11.5 concludes.

11.2 No-Arbitrage Dynamics of the Risky Asset Values

There is a traded riskless security with constant rate of return r. There are also two traded risky securities, whose market values are S_1 and S_2, respectively. Their value dynamics load asset-specific diffusive shocks and a systemic Poisson-type shock (its intensity under the objective probability measure $\mathbb{P}$ is λ):

[1]Liu et al. (2003) carry out a single-asset optimal portfolio analysis in the presence of jump-like event risk.

[2]De Donno et al. (2005) conduct a formal study of expected-utility maximization from terminal wealth in a semimartingale model with countably many assets.

[3]Merton (1971) examines asset allocation when the riskless asset follows a jump-to-ruin process.

$$\begin{aligned}
\frac{dS_1}{S_1} &= \mu_1 dt + \sigma_1 dz_1 - \eta_1 \left(dN - \lambda dt\right), \\
\frac{dS_2}{S_2} &= \mu_2 dt + \sigma_2 dz_2 - \eta_2 \left(dN - \lambda dt\right), \\
dz_1 dz_2 &= \rho dt.
\end{aligned}$$

The risky security 1 defaults in the wake of the first systemic event and may have recovery value ($0 < \eta_1 \leq 1$). The risky security 2 is hit by systemic events ($0 < \eta_2 < 1$) but survives to them. Markets are arbitrage-free with the state-price density process $\{\zeta\}$ having the dynamics[4]

$$\begin{aligned}
\frac{d\zeta}{\zeta} &= -rdt - \xi dz_\varsigma + \eta \left(dN - \lambda dt\right), \\
dz_1 dz_\varsigma &= \rho_1 dt, \\
dz_2 dz_\varsigma &= \rho_2 dt,
\end{aligned}$$

where ξ is the market price of systematic diffusive risk. I assume that systemic risk is always systematic ($\eta > 0$) and thus impacts the investment opportunity set. Indeed, by no arbitrage, the per-annum expected returns on the risky securities load the systemic-and-default risk parameters λ, η_1, and η_2 via the systematic-risk parameter η.

Proposition 1 *The no-arbitrage assumption implies that*

$$\begin{aligned}
\underbrace{\frac{1}{dt} E_t \left[\frac{dS_1}{S_1}\right]}_{=\mu_1} &= r + \xi \sigma_1 \rho_1 + \eta \eta_1 \lambda, \\
\underbrace{\frac{1}{dt} E_t \left[\frac{dS_2}{S_2}\right]}_{=\mu_2} &= r + \xi \sigma_2 \rho_2 + \eta \eta_2 \lambda.
\end{aligned}$$

Proof By no-arbitrage, the deflated value processes $\{\zeta S_1\}$ and $\{\zeta S_2\}$ must be martingales over any finite time horizon. Hence, they must be driftless:

$$E_t \left[\frac{d\left(\zeta S_i\right)}{\zeta S_i}\right] = \left(\mu_i + \eta_i \lambda\right) + \left(-r - \eta \lambda\right) + \left(-\xi \sigma_i \rho_i\right) + \left(\left(1 + \eta\right)\left(1 - \eta_i\right) - 1\right) \lambda = 0,$$

for $i = 1, 2$. □

[4]When the values of the risky securities jump at predictable dates (as at dividend dates), the state-price density has a different structure with a distinct dynamics (e.g. Battauz 2003).

11.3 The Optimal Investment Rule

The fraction of wealth allocated to the two risky securities are ϕ_1 and ϕ_2 respectively. The investor's time horizon and value function are τ and

$$J(W,\tau) = \max_{\phi_1,\phi_2} E_0\left[\frac{(W_\tau)^{1-\gamma}}{1-\gamma}\right] \quad \text{s.t.} \quad dW = W\left[\begin{array}{c} r + \phi_1\left(\frac{dS_1}{S_1} - rdt\right) + \\ \phi_2\left(\frac{dS_2}{S_2} - rdt\right) \end{array}\right],$$

respectively. The boundary condition if the first systemic event occurs immediately is

$$J(W,\tau) = H(W,\tau),$$

where

$$H(W,\tau) = \max_{\theta_2} E_0\left[\frac{(W_\tau)^{1-\gamma}}{1-\gamma}\right] \text{ with } dW = W\left[r + \theta_2\left(\frac{dS_2}{S_2} - rdt\right)\right].$$

Proposition 2 (Liu et al. 2003) *The value function at the first systemic event is*

$$H(W,\tau) = \frac{W^{1-\gamma}}{1-\gamma}e^{B(\tau)},$$

$$B(\tau) = \left(\begin{array}{c} (1-\gamma)\left(r + \theta_2^*\left(\xi\sigma_2\rho_2 + (\eta+1)\,\eta_2\lambda\right)\right) - \\ \frac{1}{2}\gamma(1-\gamma)\,\theta_2^{*2}\sigma_2^2 + \left(\left(1-\theta_2^*\eta_2\right)^{1-\gamma} - 1\right)\lambda \end{array}\right)\tau,$$

and the corresponding optimal investment rule is τ-invariant:

$$\theta_2^* = \frac{\xi\rho_2}{\gamma\sigma_2} + \frac{(\eta+1)\,\eta_2\lambda}{\gamma\sigma_2^2} - \frac{\left(1-\theta_2^*\eta_2\right)^{-\gamma}\eta_2\lambda}{\gamma\sigma_2^2}.$$

Proof The optimal investment problem is equivalent to the one considered by Liu et al. (2003) with the significant addition of the no-arbitrage assumption, which renders the speculative component of the optimal portfolio θ_2^*, $\frac{\xi\rho_2}{\gamma\sigma_2} + \frac{(\eta+1)\eta_2\lambda}{\gamma\sigma_2^2}$, dependent on the systemic-risk parameters λ and η_2. □

The optimal portfolio θ_2^* embeds a negative hedging component,

$$-\frac{\left(1-\theta_2^*\eta_2\right)^{-\gamma}\eta_2\lambda}{\gamma\sigma_2^2} < 0,$$

which is meant to create a windfall whenever the value of the risky security 2 is taken down by a systemic event. The hedging component makes sure that θ_2^* always stays below $1/\eta_2$. Notably, hedging against systemic risk is implemented even by the myopic investor ($\tau \to 0$) and by the log-utility investor ($\gamma = 1$).

While the post-default optimal portfolio θ_2^* is constant, the boundary condition at the first systemic event causes the dependence of the pre-default optimal portfolio $\left\{\phi_1^*(\tau), \phi_2^*(\tau)\right\}$ on the investment horizon τ for any investor with $\gamma \neq 1$.

Proposition 3 *The optimal investment rule* $\left\{\phi_1^*(\tau), \phi_2^*(\tau)\right\}$ *is such that*

$$\phi_1^*(\tau) = \frac{\xi(\rho_1 - \rho\rho_2)}{\gamma(1-\rho^2)\sigma_1} + \frac{\left(\eta_1 - \frac{\rho\sigma_1}{\sigma_2}\eta_2\right)(\eta+1)\lambda}{\gamma(1-\rho^2)\sigma_1^2} - \frac{\left(\eta_1 - \frac{\rho\sigma_1}{\sigma_2}\eta_2\right)\left(1 - \phi_1^*(\tau)\eta_1 - \phi_2^*(\tau)\eta_2\right)^{-\gamma} e^{-A(\tau)+B(\tau)}\lambda}{\gamma(1-\rho^2)\sigma_1^2}$$

$$\phi_2^*(\tau) = \frac{\xi(\rho_2 - \rho\rho_1)}{\gamma(1-\rho^2)\sigma_2} + \frac{\left(\eta_2 - \frac{\rho\sigma_2}{\sigma_1}\eta_1\right)(\eta+1)\lambda}{\gamma(1-\rho^2)\sigma_2^2} - \frac{\left(\eta_2 - \frac{\rho\sigma_2}{\sigma_1}\eta_1\right)\left(1 - \phi_1^*(\tau)\eta_1 - \phi_2^*(\tau)\eta_2\right)^{-\gamma} e^{-A(\tau)+B(\tau)}\lambda}{\gamma(1-\rho^2)\sigma_2^2}$$

where $A(\tau)$ *is the* τ*-dependent component of the investor's value function:*

$$J(W,\tau) = \frac{W^{1-\gamma}}{1-\gamma} e^{A(\tau)} \quad \text{with} \quad A(0) = 0.$$

Proof The Hamilton-Jacobi-Bellman equation for the investment problem[5] reads

$$0 = \max_{\phi_1,\phi_2} \left(\begin{array}{c} -A' + (1-\gamma)\begin{pmatrix} r + \phi_1(\xi\sigma_1\rho_1 + (\eta+1)\eta_1\lambda) + \\ \phi_2(\xi\sigma_2\rho_2 + (\eta+1)\eta_2\lambda) \end{pmatrix} - \\ \\ \frac{1}{2}\gamma(1-\gamma)\left(\phi_1^2\sigma_1^2 + \phi_2^2\sigma_2^2 + 2\phi_1\phi_2\sigma_1\sigma_2\rho\right) + \\ \\ \left((1-\phi_1\eta_1 - \phi_2\eta_2)^{1-\gamma} e^{-A+B} - 1\right)\lambda \end{array} \right).$$

The result follows from the first order conditions. □

[5] Alternative solution techniques for dynamic asset allocation problems include duality-based methods (e.g. Battauz et al. 2015).

When hedging against systemic risk, the investor corrects the pre-default demands for each security,

$$-\frac{\left(\eta_i-\frac{\rho\sigma_i}{\sigma_j}\eta_j\right)\left(1-\phi_i^*(\tau)\,\eta_i-\phi_j^*(\tau)\,\eta_j\right)^{-\gamma}e^{-A(\tau)+B(\tau)}\lambda}{\gamma\,(1-\rho^2)\,\sigma_i^2},\quad i=1,2,\;\; j=1,2,\;\; i\neq j,$$

to account for the normal-times correlation ρ between the returns of both securities. Importantly, the pre-default hedging demands are related to the τ-dependent ratio

$$\frac{\left(1-\phi_1^*(\tau)\,\eta_1-\phi_2^*(\tau)\,\eta_2\right)^{-\gamma}e^{B(\tau)}}{e^{A(\tau)}}$$

between the post-default marginal indirect utility of wealth and the pre-default one. The pre-default hedging demands grant that the investor's wealth at the first systemic event remains positive.

Proposition 4 *Given*

$$Y_1\left(\phi_1,\phi_1'\right)=\frac{\phi_1'}{\frac{\xi\sigma_1(\rho_1-\rho\rho_2)}{\gamma\left(1-\rho^2\right)\sigma_1^2}+\frac{1}{\gamma\left(1-\rho^2\right)}\frac{(\eta+1)\lambda}{\sigma_1^2}\left(\eta_1-\frac{\rho\sigma_1}{\sigma_2}\eta_2\right)-\phi_1}$$

and

$$Y_2\left(\phi_2,\phi_2'\right)=\frac{\phi_2'}{\frac{\xi\sigma_2(\rho_2-\rho\rho_1)}{\gamma\left(1-\rho^2\right)\sigma_2^2}+\frac{1}{\gamma\left(1-\rho^2\right)}\frac{(\eta+1)\lambda}{\sigma_2^2}\left(\eta_2-\frac{\rho\sigma_2}{\sigma_1}\eta_1\right)-\phi_2},$$

the optimal investment rule $\left\{\phi_1^*(\tau),\phi_2^*(\tau)\right\}$ is the solution to the following system of non-linear ordinary differential equations,

$$\begin{cases}
Y_1\left(\phi_1,\phi_1'\right)+\gamma\frac{\phi_1'\eta_1+\phi_2'\eta_2}{(1-\phi_1\eta_1-\phi_2\eta_2)}+B' = Z\left(\phi_1,\phi_2\right)\\
Y_1\left(\phi_1,\phi_1'\right) = Y_2\left(\phi_2,\phi_2'\right)\\
\frac{\xi\sigma_1(\rho_1-\rho\rho_2)}{\gamma\left(1-\rho^2\right)\sigma_1^2}+\frac{\left(\eta_1-\frac{\rho\sigma_1}{\sigma_2}\eta_2\right)\left(\eta+1-(1-\phi_1(0)\eta_1-\phi_2(0)\eta_2)^{-\gamma}\right)\lambda}{\gamma\left(1-\rho^2\right)\sigma_1^2} = \phi_1(0)\\
\frac{\xi\sigma_2(\rho_2-\rho\rho_1)}{\gamma\left(1-\rho^2\right)\sigma_2^2}+\frac{\left(\eta_2-\frac{\rho\sigma_2}{\sigma_1}\eta_1\right)\left(\eta+1-(1-\phi_1(0)\eta_1-\phi_2(0)\eta_2)^{-\gamma}\right)\lambda}{\gamma\left(1-\rho^2\right)\sigma_2^2} = \phi_2(0)
\end{cases},$$

where $Z(\phi_1, \phi_2)$ is quadratic in ϕ_1 and ϕ_2:

$$
\begin{aligned}
Z(\phi_1,\phi_2) = {} & (1-\gamma)\left(r + \phi_1\left(\xi\sigma_1\rho_1 + (\eta+1)\,\eta_1\lambda\right) + \phi_2\left(\xi\sigma_2\rho_2 + (\eta+1)\,\eta_2\lambda\right)\right) - \\
& \frac{1}{2}\gamma\,(1-\gamma)\left(\phi_1^2\sigma_1^2 + \phi_2^2\sigma_2^2 + 2\phi_1\phi_2\sigma_1\sigma_2\rho\right) + \\
& \frac{1}{\eta_1}\begin{pmatrix} (1-\phi_1\eta_1-\phi_2\eta_2)\left(\xi\sigma_1\rho_1 + (\eta+1)\,\eta_1\lambda\right) - \\ \gamma\,(1-\phi_1\eta_1-\phi_2\eta_2)\left(\phi_1\sigma_1^2 + \phi_2\sigma_1\sigma_2\rho\right)\end{pmatrix} - \lambda.
\end{aligned}
$$

Proof The first order conditions imply

$$
\begin{aligned}
A' &= \frac{\phi_1'}{\frac{\xi\sigma_1(\rho_1-\rho\rho_2)}{\gamma(1-\rho^2)\sigma_1^2} + \frac{1}{\gamma(1-\rho^2)}\frac{(\eta+1)\lambda}{\sigma_1^2}\left(\eta_1 - \frac{\rho\sigma_1}{\sigma_2}\eta_2\right) - \phi_1} + \gamma\frac{\phi_1'\eta_1 + \phi_2'\eta_2}{(1-\phi_1\eta_1-\phi_2\eta_2)} + B' \\
&= \frac{\phi_2'}{\frac{\xi\sigma_2(\rho_2-\rho\rho_1)}{\gamma(1-\rho^2)\sigma_2^2} + \frac{1}{\gamma(1-\rho^2)}\frac{(\eta+1)\lambda}{\sigma_2^2}\left(\eta_2 - \frac{\rho\sigma_2}{\sigma_1}\eta_1\right) - \phi_2} + \gamma\frac{\phi_1'\eta_1 + \phi_2'\eta_2}{(1-\phi_1\eta_1-\phi_2\eta_2)} + B',
\end{aligned}
$$

and

$$
\frac{1}{\eta_1}\begin{pmatrix} \begin{pmatrix} 1-\phi_1\eta_1- \\ \phi_2\eta_2 \end{pmatrix}\begin{pmatrix} \xi\sigma_1\rho_1+ \\ (\eta+1)\,\eta_1\lambda \end{pmatrix} - \\ \\ \gamma\begin{pmatrix} 1-\phi_1\eta_1- \\ \phi_2\eta_2 \end{pmatrix}\begin{pmatrix} \phi_1\sigma_1^2+ \\ \phi_2\sigma_1\sigma_2\rho \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 1-\phi_1\eta_1- \\ \phi_2\eta_2 \end{pmatrix}^{1-\gamma} e^{B-A}\lambda.
$$

□

11.4 Numerical Analysis

When looking at the U.S. dollar monthly log returns for the developed-country equity indexes for the United States, United Kingdom, Japan, Germany, Switzerland, and France over the period January 1982 to February 1997, Das and Uppal (2004) estimate that one collective jump is expected every 20 months. This implies a per-annum jump intensity level of $12/20$. More conservatively, I shrink it to $\lambda = \frac{1}{5}$, so that one systemic event is expected every 5 years. I fix the other parameter values to be $r = 0.02$, $\xi = 1$, $\eta = 2$, $\eta_1 = 0.95$, $\eta_2 = 0.4$, $\rho_1 = 0.5$, $\rho_2 = 0.5$, $\sigma_1 = 0.40$, $\sigma_2 = 0.20$, and $\rho = 0.5$. The corresponding excess expected returns on the two risky securities are

$$
\mu_1 - r = 60\% \qquad \text{and} \qquad \mu_2 - r = 28\%.
$$

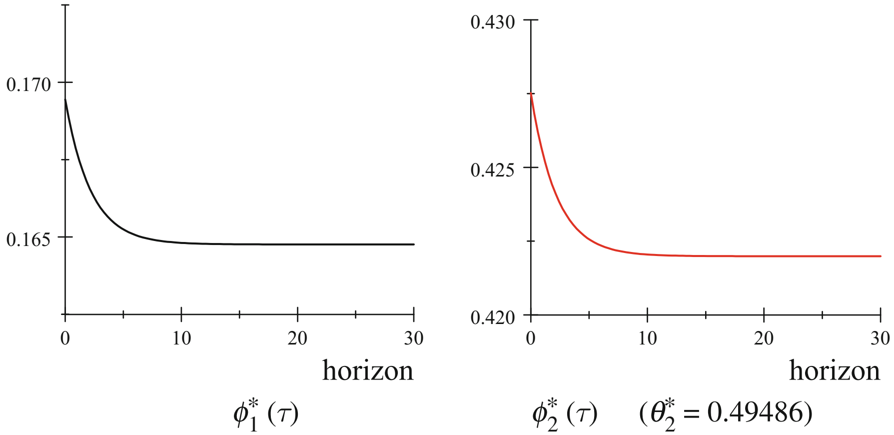


Fig. 11.1 Optimal investment versus the horizon τ (in years, $\gamma = 3$)

The pricing kernel is assumed to be highly impacted by systemic events ($\eta = 2$) and more than half of the two excess expected returns come from the systemic-risk premia ($\eta\eta_1\lambda = 38\%$ and $\eta\eta_2\lambda = 16\%$). Figure 11.1 shows the optimal investment rule plotted against the investment horizon τ for the level $\gamma = 3$ of risk aversion.

The bulk of the non-myopic hedging against systemic risk takes place at short-to-medium investment horizons. A 5-year investor subtracts about 1% of her initial wealth from the optimal myopic total exposure to risky assets, which is $\phi_1^*\ (0) + \phi_2^*\ (0) = 0.59696$. By having the lowest recovery rate ($1-\eta_1 = 5\%$), the defaultable asset is the best hedging tool against the first systemic event, that is the occurrence of default itself. In fact, the percentage difference of the 5-year single-asset allocation with respect to the myopic one is larger for the defaultable asset than for the non-defaultable asset. The optimal fraction of wealth allocated to the non-defaultable asset remains below 43% before default and leaps to more than 78% just after it (the post-default optimal portfolio is $\theta_2^* = 0.78136$).

Figures 11.2 and 11.3 show the optimal portfolio choice of more conservative investors ($\gamma = 5$ and $\gamma = 7$). On a relative basis, higher risk aversion amplifies the optimal non-myopic hedging act and confirms the defaultable asset as the best hedging tool.

11.5 Conclusions

Systemic events typically translate into a simultaneous drop in the value of many of the existing tradable assets and are often accompanied by the default of some of them. I study optimal non-myopic portfolios exposed to the joint presence of default

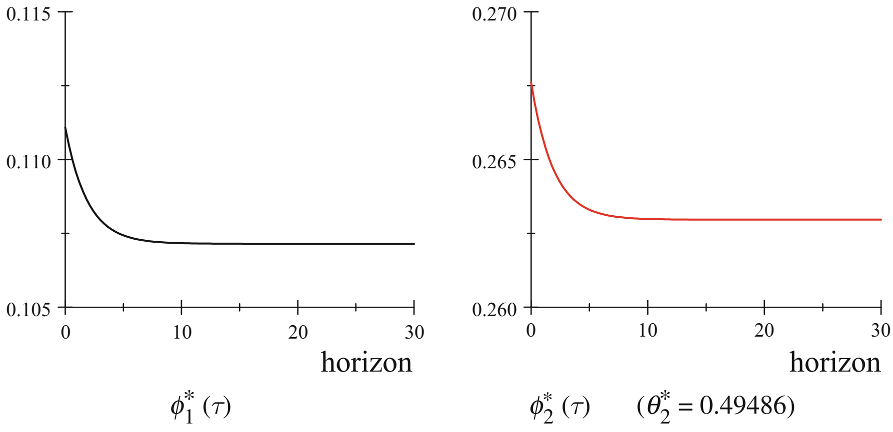


Fig. 11.2 Optimal investment versus the horizon τ (in years, $\gamma = 5$)

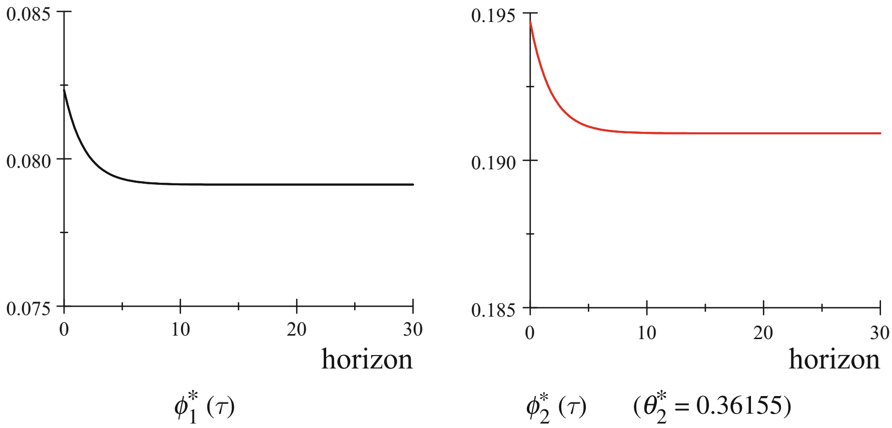


Fig. 11.3 Optimal investment versus the horizon τ (in years, $\gamma = 7$)

risk and systemic risk in no-arbitrage multiple-asset markets with a constant pre-default investment opportunity set. The assumption of no arbitrage brings balance to the optimal investment analysis by connecting excess expected returns to the parameters that express default and systemic risks. An important implication of default risk is that the ratio between the investor's post-default marginal indirect utility of wealth and the pre-default one depends on the investment horizon. Such a ratio enters the optimal intertemporal allocation of wealth. Optimal non-myopic hedging against default and systemic risks ensues.

References

A. Ang, G. Bekaert, International asset allocation with regime shifts. Rev. Financ. Stud. **15**, 1137–1187 (2002)

A. Battauz, Quadratic hedging for asset derivatives with discrete stochastic dividends. Insur. Math. Econ. **32**, 229–243 (2003)

A. Battauz, M. De Donno, A. Sbuelz, Kim and Omberg revisited: the duality approach. J. Probab. Stat. **2015**(581854), 1–6 (2015). doi:10.1155/2015/581854

N. Branger, H. Kraft, C. Meinerding, What is the impact of stock market contagion on an investor's portfolio choice? Insur. Math. Econ. **45**(1), 94–112 (2009)

N. Branger, H. Kraft, C. Meinerding, Partial information about contagion risk, self-exciting processes and portfolio optimization. J. Econ. Dyn. Control. **39**, 18–36 (2014)

A. Buraschi, P. Porchia, F. Trojani, Correlation risk and optimal portfolio choice. J. Financ. **65**(1), 393–420 (2010)

S.R. Das, R. Uppal, Systemic risk and international portfolio choice. J. Financ. **59**, 2809–2834 (2004)

M. De Donno, P. Guasoni, M. Pratelli, Super-replication and utility maximization in large financial markets. Stoch. Process. Appl. **115**(12), 2006–2022 (2005)

M. Guidolin, A. Timmermann, Asset allocation under multivariate regime switching. J. Econ. Dyn. Control. **31**, 3503–3544 (2007)

M. Guidolin, A. Timmermann, International asset allocation under regime switching, skew and kurtosis preferences. Rev. Financ. Stud. **21**, 889–935 (2008)

E. Kole, K. Koedijk, M. Verbeek, Portfolio implications of systemic crises. J. Bank. Financ. **30**, 2347–2369 (2006)

H. Kraft, M. Steffensen, Asset allocation with contagion and explicit bankruptcy procedures. J. Math. Econ. **45**, 147–167 (2009)

J. Liu, F.A. Longstaff, J. Pan, Dynamic asset allocation with event risk. J. Financ. **58**, 231–259 (2003)

R.C. Merton, Optimum consumption and portfolio rules in a continuous-time model. J. Econ. Theory **3**, 373–413 (1971)

S. Pagliarani, T. Vargiolu, Portfolio optimization in a defaultable Lévy-driven market model. OR Spectr. Financ. Optim. (2014). doi:10.1007/s00291-014-0374-7

G.W. Puopolo, Portfolio selection with transaction costs and default risk, Working paper (2015)

Chapter 12
Optimal Execution Strategy in Liquidity Framework Under Exponential Temporary Market Impact

Chiara Benazzoli and Luca Di Persio

Abstract In the present work we compute the optimal liquidation strategy for an investor who intends to entirely extinguish his position in an illiquid asset so as to minimize a criterion involving mean and variance of the strategies implementation shortfall. The market impact due to illiquidity is modeled by splitting it into two different component, namely the permanent market impact, which is assumed to be linear in the rate of trading, and the temporary market impact, which follows an exponential-type function.

Keywords Stochastic mean-variance optimization • Non-liquid markets • Non linear market impact factors • Lambert function

12.1 Introduction

We consider an optimal execution problem in a non-liquid market for a risky asset, hence allowing for an agent to influence the asset price process by participating in the market. The price variation due to the agent's actions is called market impact, and, usually, when large trades are executed, price moves in the trader's unfavorable direction, proportionally to sales volume. Therefore a common practice is to divide a large trade into many smaller ones. The main aim of this work is indeed to find the best strategy for a big sale, that is how to split it into smaller orders so as minimize the corresponding implementation cost or a cost criterion stated a priori, which may also involve risk parameters.

We solve latter problem for a model characterized by a market impact composed by two factors: the *permanent market impact* and the *temporary market impact*.

C. Benazzoli (✉)
Department of Mathematics, University of Trento, Via Sommarive, 14, 38123 Povo, Italy
e-mail: c.benazzoli@unitn.it

L. Di Persio
Department of Computer Science, University of Verona, Strada le Grazie, 15, 37134 Verona, Italy
e-mail: luca.dipersio@univr.it

G. Consigli et al. (eds.), *Handbook of Recent Advances in Commodity and Financial Modeling*, International Series in Operations Research & Management Science 257,
DOI 10.1007/978-3-319-61320-8_12

Let us recall that the permanent impact refers to the long lasting modifications of prices under the action of a given sell order, otherwise such effects are considered as temporary market impacts.

Taking into account both analytical and empirical research, see, e.g., Almgren et al. (2005), we shall consider a model given by a linear permanent market impact plus an exponential-type temporary market impact which is characterized by properly chosen parameters whose meaning will be later clarified, see Sect. 12.3.

The quest for optimal selling strategies in liquidity frameworks has become a central topic in financial mathematics during recent years. In Almgren and Chriss (1999, 2001), Almgren and Chriss consider an asset price process following an arithmetic random walk, with constant volatility over the strategy's lifetime, in the discrete-time, and an arithmetic Brownian motion (ABM), in continuous-time. In both cases, the optimal trading strategy is, by definition, the one which minimizes a linear combination of the expected cost and the variance of the cost of each strategy. Moreover in Almgren and Chriss (2001) and in Almgren and Chriss (1999), the market impact is assumed to be linear in block trades, while in Almgren (2003), it is modeled by a nonlinear function of power-law type. It is worth to mention that the latter results are based on the assumptions that the drift, the volatility of the price process and the liquidity parameters are constant over all the liquidation interval. Improvements can be achieved by modeling such parameters by stochastic processes, see, e.g., Cheridito and Sepin (2014), where the drift of the price process is forced equal to zero, or Benazzoli and Di Persio (2015), where the drift is a stochastic process. In particular in Benazzoli and Di Persio (2014), although the price process is still assumed to be an ABM, the authors replace the variance approach by considering both Value at Risk (VaR) and Expected Shortfall (ES) as risk parameters for which they exhibit related optimal execution strategies, hence obtaining more realistic results.

In Gatheral and Schied (2011), resp. in Brigo and Di Graziano (2014), the same problem outlined before is studied under the assumption that the unaffected price process is modeled as a geometric Brownian motion, resp. as a displaced diffusion process, when both components of the market impact are still assumed to be linear in the trading speed. Under latter hypothesis in Schied (2013) the authors provide a robustness property for the optimal strategies. Indeed, under a specified cost criterion, the form of the solution is independent of the unaffected price process as long as it is a square integrable martingale.

It is worth to mention that the market impact has emerged as a fundamental topic in modern electronic market. Indeed, the use of computer algorithms, and related high-frequency trading strategies, have changed a lot how transactions are currently executed. In particular the execution's speed has been modified with several implications concerning the volume size of trades. Latter scenario has been studied, e.g., in Cartea and Jaimungal (2013, 2015) and Obizhaeva and Wang (2013) from the *limit order book market* point of view. The dynamic relation between volume and returns is also studied in Llorente et al. (2002), where the authors focus on the difference between the return dynamics generated if the investors trade to hedge or to speculate.

The remainder of this paper is organized as follows. In Sect. 12.2, following Almgren and Chriss (1999) and Almgren (2003), we state the mathematical setting of the optimal trading problem we want to study. In Sect. 12.3, the originality of our approach is outlined and the optimal execution strategy is computed in terms of the Lambert W function, when the temporary market impact is modeled by an exponential-type function. Lastly, in Appendix A, the main characteristics and properties of the Lambert W function are summarized.

12.2 The Model Framework

The model is built following the framework given in Almgren and Chriss (1999). A trader holds $X \in \mathbb{R}^+$ shares of a non-liquid asset and he aims at completely liquidate his position within a fixed deadline (*fixed horizon*), $T > 0$. We divide the time interval $[0, T]$ into a finite number $N \in \mathbb{N}^+$ of subintervals of equal length $\tau := \frac{T}{N}$. Then at every discrete time $t_{n-1} = (n-1)\tau$ the trader chooses how many shares y_n to sell in the subsequent subinterval $(t_{n-1}, t_n]$. The $N-$tuple $(y_1, \ldots, y_N)$, which is called *trading list*, takes into account all the sold quantities. Notice that, since the trader sells his entire position over the whole time interval $[0, T]$, a trading list has to satisfy the liquidation constraint $\sum_{n=1}^{N} y_n = X$. By knowing the trading list, we can compute the execution strategy x, defined as a $(N+1)$-tuple $x = (X, x_1, \ldots, x_N)$ where x_n stands for the volume of shares held by the trader at time t_n, $n = 0, \ldots, N$. Since the sold quantity in the time interval $(t_{n-1}, t_n]$ matches the difference between the quantities held at the endpoints, i.e. $y_n = x_{n-1} - x_n$, then, exploiting the liquidation constraint, we have that

$$x_n = X - \sum_{k=1}^{n} y_k = \sum_{k=n+1}^{N} y_k \,,$$

which implies $x_N = 0$.

In order to model the illiquid features of the asset we split the impact due to the acting of the trader in the market, that is called market impact, into two parts, namely the permanent, resp. the temporary, component. In particular, while the temporary market impact refers to the asset price modification in the k-th time interval due to the sale occurred in the immediately preceding time interval, the permanent market impact takes into account the price variation that persists throughout the remaining trading time.

According to well established literature, see, e.g., Almgren and Chriss (2001), we assume that the unaffected price process S, that is the price per share of the asset which occurs in a *market impact-free world* or, similarly, the one we have if the trader does not participate in the market, follows an arithmetic random walk. It follows that, when the initial asset price is a known value S_0, the price per share at time t_n is given by

$$S_n = S_{n-1} + \sigma\sqrt{\tau}\xi_n \,,$$

where $\xi_1, \ldots, \xi_N$ are independent and identically distributed random variables, having zero mean and unitary variance, σ being the volatility of the asset process, which is assumed to be constant over the whole time interval $[0, T]$, as in the Black-Scholes model.

Nevertheless, the presence of liquidity effects implies that the trader does not receive the price S_t per share. Actually, the value at which the asset will be sold may be rather different. Let us underline that the price that the trader actually receives on each trade per share is called actual price (process, since it depends on time), and it will be denoted by $\tilde{S}_t$. This latter *price process*, also depending on the unaffected price as well as on the behaviour of the trader in the market, can be defined in two steps:

step 1: First we consider the permanent market impact, S_t being redefined as follows

$$S_n = S_{n-1} + \sigma\sqrt{\tau}\xi_n - \tau g(v_n) = S_0 + \sigma\sqrt{\tau}\sum_{k=1}^{n}\xi_k - \tau\sum_{k=1}^{n} g(v_k) \,, \tag{12.1}$$

where v_n is the speed of selling, i.e. it indicates the rate $\frac{y_n}{\tau}$, while the function $g(v)$ models the permanent market impact.

step 2: Then we consider the temporary component of the market impact, which is modeled by the function $h(v)$, $\tilde{S}_n$ being defined by

$$\tilde{S}_n := S_{n-1} - h(v_n) \,, \tag{12.2}$$

for $n = 1, \ldots, N$.

Exploiting (12.1) and (12.2), we can explicitly provide the difference between the two components of the market impact. In particular, at a fixed time $t_n > 0$, the actual price $\tilde{S}_n$ depends through the temporary market impact h only on the sale executed at this time, i.e. y_n, while, *vice versa*, it depends through the permanent market impact g on all the previous sold quantities $y_1, \ldots, y_n$.

The *total capture*, indicated by $G(x)$ with respect to the chosen strategy x, is nothing but the total cash received over the strategy lifetime $[0, T]$, namely

$$\begin{aligned} G(x) &:= \sum_{n=1}^{N} y_n\tilde{S}_n \\ &= \sum_{n=1}^{N} y_n S_0 + \sum_{n=1}^{N}\sum_{k=1}^{n-1}\left[(\sigma\sqrt{\tau}\xi_k - \tau g(v_k))y_n\right] - \sum_{n=1}^{N} y_n h(v_n) \end{aligned}$$

$$= XS_0 + \sigma\sqrt{\tau}\sum_{n=1}^{N} x_n\xi_n - \tau\sum_{n=1}^{N} x_n g(v_n) - \tau\sum_{n=1}^{N} v_n h(v_n)\,,$$

where the last equality follows from the relation between the trading strategy x and the related sold quantities y, as stated above. The quantity XS_0 is the *market-to-market* value of the trader's initial position, hence the difference $C(x) := XS_0 - G(x)$ is the cost due to the illiquidity. Latter quantity is often called *implementation shortfall* and it represents the ex-post measure of transaction cost. By previous assumptions on ξ_n, the cost related to a trading strategy x, i.e. $C(x)$, becomes a random variable with mean

$$\mathbb{E}[C(x)] = \tau\sum_{n=1}^{N} x_n g(v_n) + \tau\sum_{n=1}^{N} v_n h(v_n)\,, \tag{12.3}$$

and variance

$$\mathrm{Var}[C(x)] = \tau\sum_{n=1}^{N} \sigma^2 x_n^2\,. \tag{12.4}$$

In order to work in a continuous-time framework, we let the time step τ go to zero. Then a strategy is represented by a continuous function $x\colon\ [0,T] \to \mathbb{R}_0^+$ which satisfies the initial condition and the liquidation constraint if its boundary conditions are $x(0) = X$ and $x(T) = 0$. Moreover we assume that the sold quantities y_n are such that $v_n \to v(k\tau)$ when $\tau \to 0$, with $v(t) = -\dot{x}(t)$. All such strategies are called admissible and the set of all the admissible strategies will be denoted by $\mathcal{A}$. In this case, the expected value and the variance of the implementation shortfall $C(x)$, i.e. Eqs. (12.3) and (12.4), have the following finite limits:

$$\mathbb{E}[C(x)] = \int_0^T x(t)g(v(t)) + v(t)h(v(t))\,dt\,,$$

and

$$\mathrm{Var}[C(x)] = \int_0^T \sigma^2 x(t)^2\,dt\,.$$

In order to decide the optimal strategy within the set $\mathcal{A}$, we assume that the trader's goal is to find the strategy which minimizes the mean-variance cost functional U, defined as follows

$$U(x) := \mathbb{E}[C(x)] + \lambda\mathrm{Var}[C(x)]\,,$$

where λ is a positive constant. We would like to underline that the mean-variance cost criterion is one of the most popular tool used to compare different trading strategies. Indeed, it is equivalent at fixing the highest values of risk, equivalently of variance, the trader is willing to tolerate, say V^*, and then looking for the strategy that minimizes the expected cost, within all the admissible strategies with $variance \leq V^*$. It follows that the risk aversion of the trader can be efficiently modeled by the parameter λ.

In our case study, the trader's problem reads as follows.

Problem 2.1 (Minimization Problem) The objective of the trader is to find, among all the admissible strategies $\mathcal{A}$, which one minimizes the cost functional U, i.e.

$$x^* = \arg\min_{x \in \mathcal{A}} U(x) \tag{12.5}$$

where

$$U(x) = \int_0^T x(t)g(v(t)) + v(t)h(v(t)) + \lambda\sigma^2 x(t)^2 \, dt\,. \tag{12.6}$$

In order to find the optimal trading strategy x^* we argue as in Almgren (2003). First of all, the integrand function in (12.6) reads as follows

$$F(x, v) = xg(v) + vh(v) + \lambda\sigma^2 x^2\,.$$

Then, the Euler-Lagrange equation guarantees that the strategy x^* in (12.5) has to satisfy

$$F_x(x, -\dot{x}) + \frac{d}{dt}F_v(x, -\dot{x}) = 0\,,$$

that is

$$0 = F_x(x, -\dot{x}) + \dot{x}F_{vx}(x, -\dot{x}) - \ddot{x}F_{vv}(x, -\dot{x})\,.$$

Furthermore, this implies that

$$\frac{d}{dt}(F(x, -\dot{x}) + \dot{x}F_v(x, -\dot{x})) = \dot{x}\,[F_x(x, -\dot{x}) + \dot{x}F_{vx}(x, -\dot{x}) - \ddot{x}F_{vv}(x, -\dot{x})] = 0 \tag{12.7}$$

and hence, integrating both sides of Eq. (12.7) from 0 to T, it follows that the optimal strategy makes the functional $F(x, -\dot{x}) + \dot{x}F_v(x, -\dot{x})$ constant. Straightforward computations give that

$$F(x, -\dot{x}) + \dot{x}F_v(x, -\dot{x}) = x(g(-\dot{x}) + \dot{x}g'(-\dot{x})) - P(-\dot{x}) \tag{12.8}$$

where the function P in (12.8) is defined as

$$P(v) := v^2 h'(v)\,. \tag{12.9}$$

Then, if we denote by v_0 the speed at which $x(t)$ hits $x = 0$, it holds that

$$F(0, v_0) - v_0 F_v(0, v_0) = -P(v_0)\ ,$$

and therefore

$$P(-\dot{x}) - P(v_0) = x(g(-\dot{x}) + \dot{x}g'(-\dot{x})) + \lambda\sigma^2 x^2\ . \tag{12.10}$$

Remark 2.2 In order to obtain explicit solutions, we assume that the permanent impact is linear in the trading rate v, that is $g(v) = \beta v$ with β positive constant. Then, no matter the strategy the trader follows, we have

$$x(g(-\dot{x}) + \dot{x}g'(-\dot{x})) = x(-\beta\dot{x} + \beta\dot{x}) = 0\ ,$$

and then the condition stated in Eq. (12.10) reduces to

$$P(-\dot{x}) = \lambda\sigma^2 x^2 + P(v_0)\,.$$

By separation of variables, and assuming that P^{-1} is well defined and $P^{-1}(\lambda\sigma^2 x^2 + P(v_0)) \neq 0$ for all $t \in [0, T]$, we obtain that the optimal strategy solves the following equation

$$\int_{x(t)}^{X} \frac{1}{P^{-1}(\lambda\sigma^2 x^2 + P(v_0))}\, dx = t\,. \tag{12.11}$$

12.3 Exponential Market Impact Function

In what follows we still consider a linear permanent market impact $g(v) = \beta v$ with $\beta > 0$, and we specify the temporary market impact as an exponential function, namely we assume

$$h(v) := \begin{cases} \gamma e^{-\frac{\theta}{v}} & \text{for} \quad v > 0, \\ 0 & \text{for} \quad v = 0, \end{cases} \tag{12.12}$$

where γ and θ are strictly positive constants. The reason why h is defined only for positive value of the trading rate will be clarified later. Through the parameters γ and θ we can control the shape of the temporary market impact. Notice that the function $h(v)$ is strictly increasing in its domain, it is convex on the set $\left[0, \frac{\theta}{2}\right]$ and concave for $v \geq \frac{\theta}{2}$. Experimental analysis confirms the concavity of the temporary impact function, see Almgren et al. (2005), Obizhaeva (2012) and reference therein. Nevertheless we choose to allow it to be convex for small values of v. In fact such values are difficult to estimate, since they correspond to small change in the price, moreover the system may be extremely fragile around a critical point.

Under previous assumptions, and without fixing a deadline T, we can explicitly compute the optimal trading strategy in the case when the set of admissible strategies is narrowed by considering only those of pure selling type. We recall that a trading strategy is called *pure selling strategy* if its rate process is strictly positive, namely if the strategy itself is strictly decreasing. From now on, a strategy x will be admissible if, besides satisfying the conditions mentioned above, it is of pure selling type.

Theorem 3.1 *Let us assume that no deadline is exogenously imposed on the sale. If the permanent market impact is linear in the trading rate and the temporary market impact is given as in* (12.12)*, the optimal solution among all the admissible pure selling strategies of the minimization Problem* 2.1 *is given by*

$$x^*(t) = \begin{cases} \dfrac{\exp\left\{ W_{-1}\left(\dfrac{\kappa\left(\frac{\theta}{2}t + X(\ln[\kappa X] - 1)\right)}{e} \right) + 1 \right\}}{\kappa} & \text{for} \quad t < \frac{2}{\theta} X \ln\left[\frac{e}{\kappa X}\right] \\ 0 & \text{for} \quad t \geq \frac{2}{\theta} X \ln\left[\frac{e}{\kappa X}\right] \end{cases} \tag{12.13}$$

where W is the Lambert W function and $\kappa = \sqrt{\frac{\lambda\sigma^2}{\gamma\theta}}$, provided that

$$\kappa X < 1 . \tag{12.14}$$

Proof According to the considerations outlined in Sect. 12.2, the optimal trading strategy we are looking for satisfies Eq. (12.11). Under the assumptions of the theorem, the function P, as defined in (12.9), becomes

$$P(v) = v^2 \frac{d}{dv} e^{-\gamma \frac{\theta}{v}} = \gamma\theta e^{-\frac{\theta}{v}} ,$$

which has as inverse function $P^{-1}(v) = -\frac{\theta}{\ln[\frac{v}{\gamma\theta}]}$, which is only defined for $v > 0$. Therefore a necessary condition for the well-posedness of problem (12.11) is indeed to consider strategies with strictly positive rate process, which implies to consider only sell programs. Therefore problem (12.11) turns out to be

$$\frac{1}{\theta} \int_{x(t)}^{X} - \ln\left[\frac{\lambda\sigma^2 x^2 + \gamma e^{-\frac{\theta}{v_0}}}{\gamma\theta} \right] dx = t . \tag{12.15}$$

If, as in this case, no time horizon is exogenously imposed then we obtain the longest possible liquidation time, denoted in the following by T, by setting $v_0 = 0$, and therefore the problem stated in (12.15) reduces to

$$\int_{x(t)}^{X} -\ln\left[\frac{\lambda\sigma^2 x^2}{\gamma\theta}\right] dx = \theta t \,. \tag{12.16}$$

In order to the problem be well-posed, the candidate solution has to satisfy the constraint $P^{-1}(\kappa^2 x(t)^2) \neq 0$ for all $t \in [0, T]$, i.e. $-\frac{1}{\ln[\kappa^2 x(t)^2]} \neq 0$ for all $t \in [0, T]$, then the optimal execution strategy must satisfy one, and only one, of the following conditions

$$\kappa x(t) < 1 \quad \forall t \in [0, T] \quad \text{or} \quad \kappa x(t) > 1 \quad \forall t \in [0, T] \,.$$

Since at the final time T the strategy's value is $x(T) = 0$, i.e. $\kappa x(T) < 1$, the optimal solution x^* can only meet the first constraint. Latter condition is verified since each admissible trading strategy $x \in \mathcal{A}$ is decreasing with $\kappa X < 1$ at the initial time as required by the theorem. It can be seen that this condition is a constraint on the model's parameters, indeed it reads as $\frac{\lambda\sigma^2 X^2}{\gamma\theta} < 1$. Equation (12.16) implies that the quantity

$$\int_{x(t)}^{X} -\ln\left[\frac{\lambda\sigma^2 x^2}{\gamma\theta}\right] dx = 2x(t)\left(\ln\left[\frac{\sqrt{\lambda}\sigma x(t)}{\sqrt{\gamma\theta}}\right] - 1\right) - 2X\left(\ln\left[\frac{\sqrt{\lambda}\sigma X}{\sqrt{\gamma\theta}}\right] - 1\right)$$

is equal to θt, and therefore the optimal strategy fulfills

$$x(t)(\ln[\kappa x(t)] - 1) = \frac{\theta}{2} t + X(\ln[\kappa x(t)] - 1) \,,$$

that can be rewritten as

$$x(t)\ln\left[\frac{\kappa x(t)}{e}\right] = \frac{\theta}{2} t + X(\ln[\kappa X] - 1) \,. \tag{12.17}$$

Equation (12.17) has two solutions for each $t < \frac{2}{\theta} X \ln\left[\frac{e}{\kappa X}\right]$. Nevertheless since we have assumed to perform a pure selling strategy and then x is a continuous and decreasing function, there exists a unique trading strategy which satisfies Eq. (12.17), namely (12.13). Notice that the optimal strategy reaches $x^* = 0$ in a finite time $T = \frac{2}{\theta} X \ln\left[\frac{\sqrt{\gamma\theta}}{\sqrt{\lambda}\sigma X} e\right]$.
See Appendix A for further details on the Lambert W function.

Remark 3.2 We want to directly verify that the optimal execution strategy stated in Theorem 3.1 satisfies the initial condition $x(0) = X$. By definition the initial value of the optimal strategy is

$$x(0) = \frac{e^{W_{-1}\left(\frac{\kappa X \sigma}{e} \ln\left[\frac{\kappa X}{e}\right]\right)+1}}{\kappa} \,,$$

and since W_{-1} is upper bounded by -1, then the initial value $x(0)$ belongs to the interval $\left[0, \frac{1}{\kappa}\right]$. Moreover, by manipulating the previous equation, we have that $x(0)$ solves

$$\frac{\kappa x(0)}{e} = e^{W_{-1}\left(\frac{\kappa X}{e}\ln\left[\frac{\kappa X}{e}\right]\right)} .$$

Hence, taking the logarithm of both sides and using the definition of the Lambert W function, we obtain the equality

$$\frac{\kappa x(0)}{e}\ln\left[\frac{\kappa x(0)}{e}\right] = \frac{\kappa X}{e}\ln\left[\frac{\kappa X}{e}\right] ,$$

which is verified by $x(0) = X$. In fact this is the unique solution of the latter equation, since the function $\frac{\kappa y}{e}\ln\left[\frac{\kappa y}{e}\right]$ is strictly decreasing in the interval $\left[0, \frac{1}{\kappa}\right]$.

12.3.1 Evaluation of W_{-1}

Even if the Lambert W function can not be expressed in terms of elementary functions, we want to describe its behaviour in order to sketch the optimal trading strategy (12.13). Since the Lambert W function is defined by mean of an inverse relation, arbitrary-precision evaluations can be obtained by iterative root-finding methods. Given a value z, its corresponding value $w = W(z)$ satisfies $we^w = z$, that is the root of the function $f(w) = we^w - z$. Notice that since the Lambert W function $W(z)$ is bi-valued in $\left(-\frac{1}{e}, 0\right)$, we have to take into account that we will find two solutions: the one that is greater than -1 is the value of the so called principal branch $W_0(z)$, while the second real branch, the lower branch, is indeed $W_{-1}(z)$.

Several numerical methods for the root finding problem have been developed, which differ each other for complexity of implementation, conditions and rate of convergence. A natural choice in our setting is to use the third-order Halley's method which starts with an initial guess w_0 for the root, and then performs the following iteration scheme

$$w_{n+1} = w_n - \frac{2f(w_n)f'(w_n)}{2(f'(w_n))^2 - f(w_n)f''(w_n)} = w_n - \frac{(w_n e^{w_n} - z)}{e^{w_n}(w_n + 1) - \frac{(w_n e^{w_n} - z)(w_n + 2)}{2(w+1)}} ,$$

which converges to the desired value.

Figure 12.1 shows the behaviour of the optimal solution x^*, see (12.13), for different values of the parameter θ, the other parameters being fixed as in Table 12.1. Notice that Theorem 3.1 applies, in fact the parameters always satisfy condition (12.14), i.e. $\frac{\lambda\sigma^2}{\gamma\theta}X < 1$.

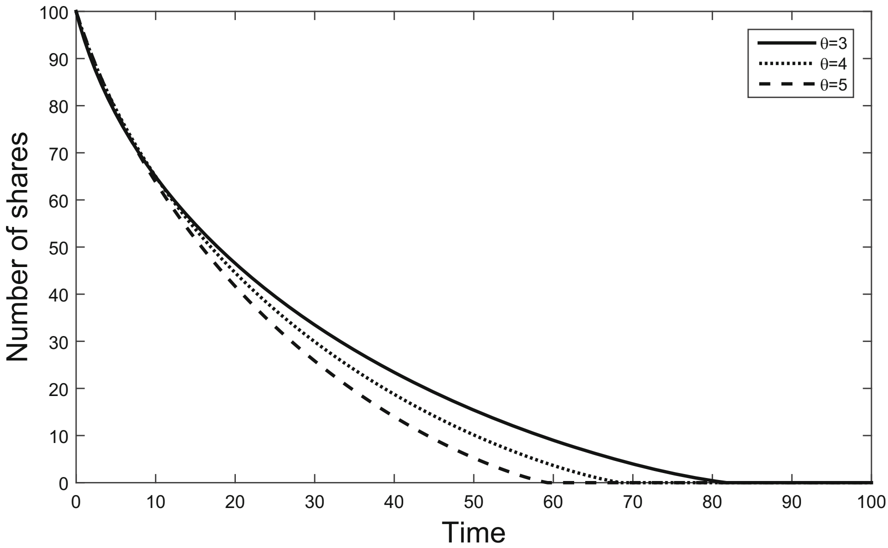


Fig. 12.1 Optimal solution strategy

Table 12.1 Parameters' value

Parameter	Value
X	100
λ	5
σ	0.02
γ	10.5

By the definition of function h in (12.12), it can be seen that, fixed a sales volume, the effect on the price due to the temporary market impact is lower for a higher value of θ. This means in particular that when θ is higher the decrease in price is smaller and therefore the trader sells the illiquid asset faster.

12.4 Conclusion

We have found the explicit optimal execution strategy that minimizes a criterion containing the expected cost and the variance of the implementation shortfall for trading a non-liquid asset when the market impact is model by an exponential-type function. In doing so, we extend the case studies by considering different nonlinear impact functions from the ones introduced in Almgren (2003).

We have also provided the optimal execution strategy under reasonable assumptions, in particular exploiting the Lambert W function. In particular, it is possible to numerical treat latter result in order to obtain the solution behavior as well as its intrinsic properties.

A Lambert W Function

In this Section, we recall the main characteristics and properties of the Lambert W function. See Corless et al. (1996) for more details. The Lambert W function, also called the omega function, is defined to be the function satisfying $z = W(z)e^{W(z)}$, that is the inverse function of $f(w) = we^w$, which is not injective, hence the relation W is multivariate. In particular, if x is real $W(x)$ is double-valued on $\left(-\frac{1}{e}, 0\right)$. Then the Lambert W function has two real branches with a branching point located at $(-e^{-1}, -1)$. Indeed if we consider W under the constraint $W \geq -1$ or $W \leq -1$, they are two well defined real valued functions. The branch satisfying $W \geq 1$ is called *principal branch* and it is denoted by W_0, or just W, if no ambiguity exists, while the branch satisfying $W \leq -1$, the *lower branch*, is denoted by W_{-1}. The Lambert W function has the special values $W(-e^{-1}) = -1$, $W(0) = 0$, and $W(1) \simeq 0.567143$, called the *omega constant*, that satisfies $\exp(-W(1)) = W(1)$, that is $\ln\left[\frac{1}{W(1)}\right] = W(1)$.

A.1 *Taylor Series for* $-\frac{1}{e} < z < 0$

The Lambert W function $W_{-1}(z)$ is upper-bounded and infinitely differentiable in $(-\frac{1}{e}, 0) \in \mathbb{R}$. By differentiating the defining expression $z = W(z)e^{W(z)}$, it follows $1 = W'(z)e^{W(z)} + W(z)W'(z)e^{W(z)}$, and then the first derivative of W turns out to be

$$W'(z) = \frac{1}{e^{W(z)}(1 + W(z))}$$

provided that $W(z) \neq -1$, i.e. $z \neq -\frac{1}{e}$ or equivalently $W'(z) = \frac{W(z)}{z(1+W(z))}$, with the additional condition $W(z) \neq 0$, i.e. $z \neq 0$. The nth derivative of W is

$$\frac{d^n W(x)}{dx^n} = \frac{e^{-nW(x)} P_n(W(x))}{(1 + W(x))^{2n-1}} \;, \text{ for } \quad n \geq 1 \;, \tag{A.18}$$

where the polynomials $P_n(w)$ are defined by the recurrence relation

$$P_{n+1}(w) = (1 - nw - 3n)P_n(w) + (1 + w)P'(w) \quad \text{for} \quad n \geq 2$$

and the initial polynomial $P_1(w) = 1$. Indeed

$$\begin{aligned}
\frac{d^{n+1}W(z)}{dx^{n+1}} &= \frac{d}{dz}\frac{e^{-nW(z)}P_n(W(z))}{(1+W(z))^{2n-1}} \\
&= \frac{[-nW'(z)e^{-nW(z)}P_n(W(z)) + e^{-nW(z)}P_n'(W(z))W'(z)](1+W(z))^{2n-1}}{(1+W(z))^{4n-2}} \\
&\quad - \frac{(2n-1)(1+W(z))^{2n-2}W'(z)e^{-nW(z)}P_n(W(z))}{(1+W(z))^{4n-2}} \\
&= \frac{e^{-(n+1)W(z)}[(1+W(z))P_n'(W(z)) + (-3m - nW(z) + 1)P_n(W(z))]}{(1+W(z))^{2n+1}}\,.
\end{aligned}$$

Then for any z_0 and z in the domain $(-\frac{1}{e}, 0)$ we can write the Taylor series for the function W_{-1} as

$$W_{-1}(z) = W_{-1}(z_0) + \sum_{n=1}^{\infty}\frac{1}{n!}W_{-1}^{(n)}(z_0)(z-z_0)^n\,.$$

Notice that since the nth derivative of W_1 in z_0, i.e. $W_1^{(n)}(z_0)$, can be computed just by knowing $W_1(z_0)$, see (A.18), it is enough to estimate the function W_{-1} in z_0 to know also the derivatives values.

A.2 Series Expansions About the Branch Point $z = -\frac{1}{e}$

For a fixed value $z \in \left[-\frac{1}{e}, 0\right)$, let us consider the point $p = -\sqrt{2(ez+1)}$, which is such that $ez = \frac{p^2}{2} - 1$. Then

$$W(z)e^{W(z)} = z \implies W(z)e^{z+1} = \frac{p^2}{2} - 1\,.$$

By expanding the exponential function in power of $W(z)+1$ we have that

$$\frac{p^2}{2} - 1 = W\sum_{k=0}^{\infty}\frac{(W(z)+1)^k}{k!} = -1 + \sum_{k=1}^{\infty}\left(\frac{1}{(k-1)!} - \frac{1}{k!}\right)(1+W(z))^k\,,$$

then we have $W_{-1}(z) = \sum_{k=0}^{\infty}\mu_k p^k$, where $\mu_k = \frac{k-1}{k+1}\left(\frac{\mu_{k-2}}{2} + \frac{\alpha_{k-2}}{4}\right) - \frac{\alpha_k}{2} - \frac{\mu_{k-1}}{k+1}$, $\alpha_k = \sum_{j=2}^{k-1}\mu_j\mu_{k+1-j}$, with $\mu_0 = -1$, $\mu_1 = 1$, $\alpha_0 = 2$, $\alpha_1 = -1$, and the series converges in the whole domain of existence of W_{-1}. For details, see e.g. Chapeau-Blondeau and Monir (2002).

A.3 Asymptotic Series for $z < 0$

A real-valued asymptotic series can be found when $z \to 0^-$. Indeed, by using the Lagrange inversion theorem, it can be found

$$W_{-1}(z) = \ln[-z] - \ln[-\ln(z)] + \sum_{k=0}^{\infty}\sum_{m=0}^{\infty} C_{km}(\ln[-z])^{-(k+m)}(\ln[-\ln[-z]])^m ,$$

with

$$C_{km} = \frac{(-1)^k S(k+m, k+1)}{m!}$$

where $S(k+m, k+1)$ is a non-negative Stirling number of the first kind. They count the number of permutations of n elements with k disjoint cycles and also arise as coefficients of the rising factorial

$$(x)^{(n)} = x(x+1)\cdots(x+n-1) = \sum_{m=0}^{n} S(n,m)x^m .$$

Moreover they are computable via the recursive formula

$$S(n,m) = S(n-1, m-1) + (n-1)S(n-1, m) \quad n > 1 .$$

References

R.F. Almgren, Optimal execution with nonlinear impact functions and trading-enhanced risk. Appl. Math. Financ. **10**(1), 1–18 (2003)

R. Almgren, N. Chriss, Value under liquidation. Risk **12**(12), 61–63 (1999)

R. Almgren, N. Chriss, Optimal execution of portfolio transactions. J. Risk **3**, 5–40 (2001)

R. Almgren et al., Direct estimation of equity market impact. Risk **18**(7), 58–62 (2005)

C. Benazzoli, L. Di Persio, Optimal execution strategy under arithmetic Brownian motion with VaR and ES as risk parameters. Int. J. Appl. Math. **27**(2), 155–162 (2014)

C. Benazzoli, L. Di Persio, Optimal execution strategy in liquidity framework. Appl. Math. Financ. (2015, Submitted)

D. Brigo, G. Di Graziano, Optimal execution comparison across risks and dynamics, with solutions for displaced diffusions. J. Financ. Eng. **1** (2014). http://www.worldscientific.com/doi/abs/10.1142/S2345768614500184

Á. Cartea, J. Sebastian, Modelling asset prices for algorithmic and high-frequency trading. Appl. Math. Finance **20**(6), 512–547 (2013)

Á. Cartea, J. Sebastian, Order-flow and liquidity provision (2015). Available at SSRN 2553154

F. Chapeau-Blondeau, A. Monir, Numerical evaluation of the Lambert W function and application to generation of generalized Gaussian noise with exponent 1/2. IEEE Trans. Signal Process. **50**(9), 2160–2165 (2002)

P. Cheridito, T. Sepin, Optimal execution under stochastic volatility and liquidity. **21**(4), 342–362 (2014)

R.M. Corless et al., On the Lambert W function. Adv. Comput. Math. **5**(1), 329–359 (1996)

J. Gatheral, A. Schied, Optimal trade execution under Geometric Brownian Motion in the Almgren and Chriss framework. Int. J. Theor. Appl. Finance **14**(03), 353–368 (2011)

G. Llorente et al., Dynamic volume-return relation of individual stocks. Rev. Financ. Stud. **15**(4), 1005–1047 (2002)

A.A. Obizhaeva, Liquidity estimates and selection bias. Technical report, University of Maryland working paper, 2012

A.A. Obizhaeva, J. Wang, Optimal trading strategy and supply/demand dynamics. J. Financ. Mark. **16**(1), 1–32 (2013)

A. Schied, Robust strategies for optimal order execution in the Almgren- Chriss framework. Appl. Math. Finance **20**(3), 264–286 (2013)

Chapter 13
Optimal Multistage Defined-Benefit Pension Fund Management

Giorgio Consigli, Vittorio Moriggia, Elena Benincasa, Giacomo Landoni, Filomena Petronio, Sebastiano Vitali, Massimo di Tria, Mario Skoric, and Angelo Uristani

Abstract We present an asset-liability management (ALM) model designed to support optimal strategic planning by a defined benefit (DB) occupational pension fund (PF) manager. PF ALM problems are by nature long-term decision problems with stochastic elements affecting both assets and liabilities. Increasingly PFs operating in the second pillar of modern pension systems are subject to mark-to-market accounting standards and constrained to monitor their risk capital exposure over time. The ALM problem is formulated as a multi-stage stochastic program (MSP) with an underlying scenario tree structure in which decision stages are combined with non-decision annual stages aimed at mapping carefully the evolution of PF's liabilities. We present a case-study of an underfunded PF with an initial liquidity shortage and show how a dynamic policy, relying on a set of specific decision criteria, is able to gain a long-term equilibrium solvency condition over a 20 year horizon.

Keywords Pension fund management • Multistage stochastic programming • Scenario tree • Solvency ratio • Defined benefits

13.1 Introduction

Most pension systems in OECD countries rely on three pillars: a state-controlled security system which represents the fundamental instrument for welfare policies, a sector-specific complementary occupational pension pillar open to corporations' and sectors' employees and, finally, individual retirement contracts that can be

G. Consigli (✉) • V. Moriggia • E. Benincasa • G. Landoni • F. Petronio • S. Vitali
Department of Management, Economics and Quantitative Methods, University of Bergamo, Bergamo, Italy
e-mail: giorgio.consigli@unibg.it

M. di Tria • M. Skoric • A. Uristani
Allianz Investment Management, Allianz Group, Munich, Germany

G. Consigli et al. (eds.), *Handbook of Recent Advances in Commodity and Financial Modeling*, International Series in Operations Research & Management Science 257,
DOI 10.1007/978-3-319-61320-8_13

agreed between pension funds, life insurers and individuals to provide deferred income during retirement. The first pillar is by definition public while the second and third pillars are mainly private. Third pillar's individual ALM problems and associated modeling and economic issues have attracted significant interest in recent years and find a natural formulation as long-term consumption-investment problems over a life-cycle (Medova et al., 2008; Consiglio et al., 2007; Consigli et al., 2012b; Konicz et al., 2015). In this chapter we analyse instead the key modeling and methodological issues needed to formulate and solve a second pillar occupational PF strategic asset allocation problem relying on a multistage stochastic program (MSP) with liabilities generated by a defined benefit (DB) scheme. ALM developments for pension funds have a relevant record starting from early works (Consigli and Dempster, 1998; Ziemba et al., 1998; Pflug and Świetanowski, 1999), then increasingly relying on real-world case-studies (Mulvey et al., 2006, 2007, 2008; Geyer and Ziemba, 2008) and again, more recently, exploring alternative modeling and optimization approaches (Aro and Pennanen, 2016; Pachamanova et al., 2016; Consigli et al., 2016). A focus on the differences between DB and defined contribution (DC) schemes can be traced in Dert (1998) and Consiglio et al. (2015).

PF ALM theory originally developed from the challenges associated with the modeling of liability streams typically extended over several decades. Those streams carry relevant risks to the PF manager. On the liability side, indeed, a DB PF manager faces mainly three relevant risk sources: inflation risk, interest rate risk and longevity risk (Aro and Pennanen, 2016; Consigli et al., 2016). In this chapter we consider a pension plan decumulation phase in which pensions are revaluated according to the previous year recorded Consumer Price Index (CPI) annual variation. The value of the pension fund liability, or DB obligation (DBO), is then computed as the discounted value of all expected pension payments and will depend on the evolution of the yield curve over the decision horizon. Such exposure is typically compensated for hedging purposes by fixed-income holdings in the asset portfolio. We assume that the PF manager, before any indexation, is informed by the actuarial division of future contributions and pension payments over a very long horizon. Longevity risk comes into the picture because upon determination of individuals' benefits through an annuity, the PF manager needs to assume a future life length that might underestimate passive members' actual future life duration (Aro and Pennanen, 2016; Shaw, 2007). Such phenomenon is attracting increasing interest by actuarial studies. In this chapter we assume that projected pension flows' estimates do already take into account such risk.

A PF manager primary concern is to preserve a sufficient funding condition: as liabilities evolve over time along random paths, the assets' market value is expected at least to match their value. The *funding ratio* (FR) is defined as the ratio of the asset portfolio to the DBO value. A FR close to 1 suggests proper funding of all future inflation-adjusted pension payments. A FR below 1 will indicate an under-funding condition. Notice that a FR below 1 doesn't necessarily imply a liquidity deficit; it does however describe a weak financial condition. The joint dependence on interest rates and inflation dynamics of assets and liabilities underlines the importance of a dynamic ALM approach to the problem and an effective risk assessment and monitoring (Pachamanova et al., 2016; Consigli et al., 2016).

We present in this chapter a novel methodological approach, inspired by Dempster et al. (2007), still in the stream of advanced applications of MSP with recourse, in which decision nodes are a subset of nodes along a scenario tree accommodating annual cash flows and obligations within what we refer to as intermediate nodes. We are able in this way to monitor accurately the cash condition of the PF and avoid anticipative portfolio rebalancing decisions. The key motivation of this work is related to the evidence that a careful representation and discrete formulation of a dynamic PF ALM model, including a liability-driven component and an innovative set of decision criteria, can consistently bring a maybe severely underfunded PF into a long-term funding equilibrium. This is what we believe will define our contribution in the first place. The model development and the definition of a case-study in cooperation with industry colleagues are also worth sharing.

The chapter evolves from Sect. 13.2 where we introduce the key elements of the ALM model and specify the scenario tree convention supporting the MSP formulation. The problem mathematical instance is introduced progressively in the first section and over the following sections to clarify the role played by each set of equations in the economy of a PF. In Sect. 13.3 the liability model is introduced. In Sect. 13.4 the risk exposure of the PF to market fluctuations is analyzed before focusing in Sect. 13.5 on funding conditions and long term PF solvency. In Sect. 13.6 we present a case study application and in Sect. 13.7 we conclude.

13.2 DB Pension Fund Management

We consider a company offering a complementary DB pension scheme to its employees: the PF collects contributions from the PF's sponsor, and marginally from the employees, and pays benefits. A DB plan ensures benefits based on the employee's remuneration at or near the retirement date. The adjustment of pension benefits to account for changes in costs and living standards is reflected in the fund's indexation rule. The level of indexation has an important impact on the income of retirees and the pension accrual of active members in pension plans.

Under a DB scheme, the PF manager faces several risk sources: core to this decision process is the need to keep the FR sufficiently high and monitor its evolution over time. The FR definition differs from the funding gap which is computed as the difference between liability and assets and coincides with the concept of net DBO. A negative net DBO thus coincides with a surplus condition of the fund.

A PF DBO is in general naturally affected by the ratio of active to passive members and a deteriorating population ratio in several economic sectors has been recorded in recent years. Furthermore survival probabilities have been increasing leading to a surge of longevity risk (Aro and Pennanen, 2016).

From a financial viewpoint, furthermore, since the 2008 global financial crisis, interest rates have materially decreased in most OECD economies, reducing the liquidity buffer generated historically by significant holdings of treasury bonds by PF managers and increasing their liabilities, ceteris paribus. Both phenomena motivated over the last decade or so an increasing role of less liquid and relatively risky portfolio strategies throughout the PF industry (OECD, 2011). In this chapter, we rely on an exogenous liability stream and focus on the definition of an optimal portfolio policy, when the population of active members is assumed not to increase (closed population model). The fund's contributions are generated by the employees, as a salary portion, and the employer, or sponsor of the pension fund. Further to such commitment, the sponsor has an obligation to fill up the PF resources in case of prolonged underfunding conditions. The PF wishes to minimize the injection into the fund of those unexpected and extraordinary contributions, comparable to undesirable recapitalization decisions. Throughout the chapter the PF manager represents the decision maker of the ALM problem.

The need of an effective portfolio policy over a long-term horizon is a strong motivation for the adoption of a dynamic approach: to preserve sufficient realism in the problem description we propose in this chapter a specific scenario tree structure, based on a distinction between decision nodes and so-called intermediate nodes, where pension payments are recorded. Such formulation makes the mapping of pension payments on an annual basis possible and allows the definition of an optimal investment strategy over a 20-year horizon with non-homogeneous time stages. The formulation of the PF ALM problem as a multistage stochastic program is shown to provide an effective way to manage the PF risk exposure and regain a stable funding condition at the end of the planning horizon.

We assume an extended investment universe including Treasury indices with different maturity, corporate indices, real estate investments, inflation-linked treasuries (TIPS) also spanning several maturities and equity. Even if indices are known to be almost constant-maturity financial benchmarks, they are here treated as investment opportunities carrying a time to maturity. Accordingly fixed income assets will generate both interests during the holding period and capital at current market prices when expiring.

The company's actuarial division is responsible for determining the expected evolution of the PF's liabilities, treated as exogenous to the definition of an optimal investment plan. In this analysis the population of active members is assumed closed and benefit cashflows are first inflation-adjusted and then discounted at a prevailing term structure of interest rates forward in time. As a result both net pension payments and the fund's DBO are considered exogenous and scenario dependent.

We adopt a simple multivariate Gaussian model to derive the random evolution of price dynamics and interest rates as well as dividend payments (see the estimated means, variances and covariances in appendix). The PF manager seeks an optimal policy based on a combination of decision criteria, including PF liquidity, duration-matching between assets and liabilities, risk-adjusted performance and sponsor's extraordinary contributions. Further at the end of the planning horizon the PF manager seeks a minimal, ideally null net DBO. A null net DBO, corresponding to a unit FR, is regarded as a Fund's equilibrium condition.

To exemplify a PF risky condition, assume for instance that a severe market shock will impact the PF asset portfolio causing a sudden decrease of its market value and to mitigate systemic risk, monetary authorities will decide a strong reduction of interest rates. In case of heavy mismatching between assets' and liabilities' duration, such intervention would add up to the market shock, inducing a further decrease of the FR and a deterioration of funding conditions. Such nightmare scenario has indeed occurred at several points of recent financial history and has deeply affected the pension industry, leading to more rigorous risk management approaches.

From a modeling viewpoint the MSP formulation leads to the definition of the optimal ALM strategy as a decision tree process describing the sequential interplay between decisions and random events. The beginning of the decision horizon coincides with the current time $t = 0$, while T will denote the end of the decision horizon. A reference period defined as 0^-, prior to 0, is introduced in the model to define an input portfolio from which the optimal time 0 portfolio will be determined. The extended time horizon is represented by $\widehat{\mathcal{T}} = \mathcal{T} \cup 0^-$, $t \in \widehat{\mathcal{T}}$. At the last stage, no investment decisions are allowed, while cash flows due to expiring bonds and income payments will be accounted for.

We denote with $T_d = \{1, 2, 3, 5, 10, 20\} \in \mathcal{T}$ a set of decision times expressed in years. Time 0 is associated with the so-called here and now decision (or H&N) and the subsequent stages have associated all scenario dependent decisions. In addition we consider $T_{int} = \mathcal{T} \backslash \{T_d\} = \{4, 6, 7, 8, 9, 11, \ldots, 19\}$ to denote a set of intermediate years; therefore $\mathcal{T} = T_{int} \cup T_d$. At intermediate times, income returns and the assets expiring provide random cash inflows to pay current pensions.

Random dynamics are modeled through discrete tree processes defined in an appropriate probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (Dupačová et al., 2001; Dempster et al., 2011; Consigli et al., 2012a). Following the model notation in Consigli et al. (2012b), nodes along the tree, for $t \in \mathcal{T}$, are denoted by $n \in \mathcal{N}_t$; we distinguish between decision nodes $\mathcal{N}_t^d$ and intermediate nodes $\mathcal{N}_t^{int}$. Again $\mathcal{N}_t = \mathcal{N}_t^d \cup \mathcal{N}_t^{int}$. For $t = 0$ the root node (associated with the partition $\mathcal{N}_0 = \{\Omega, \emptyset\}$, corresponding to the entire probability space) is labeled $n = 0$. For $t > 0$ every $n \in \mathcal{N}_t$ has a unique ancestor n^- and, for $t < T$, a non-empty set of children nodes n^+. We indicate with $m \in \mathcal{C}_n$ the nodes in the sub-tree originating from node n. We denote with t_n the time associated with node n.

The set of all predecessors of node n: $n^-, n^{--}, \ldots, 0$ is denoted by $\mathcal{P}_n$. We distinguish between ancestor decision nodes $\mathcal{P}_n^d$, i.e. $n^{d-}, n^{d--}, \ldots, 0^-$, and ancestor intermediate nodes $\mathcal{P}_n^{int}$, i.e. $n^{int-}, n^{int--}, \ldots$, and we define $\widehat{\mathcal{P}} = \mathcal{P}_n^d \cup 0^-$.

The scenario tree conditional nodal structure defines the space $(\Omega, \mathcal{F}, \mathbb{P})$ for the problem. We define the probability distribution $\mathbb{P}$ on the leaf nodes ($n \in \mathcal{N}_T$) of the scenario tree so that $\sum_{n \in \mathcal{N}_T} p_n = 1$. One scenario is identified by a sequence of nodes from the root node to one of the leaf nodes, whose cardinality will coincide with the number of scenarios. Scenarios $S = N_T$ are sample paths originated from the root node defined at time 0 to the leaf nodes at T, determine the stochastic nature

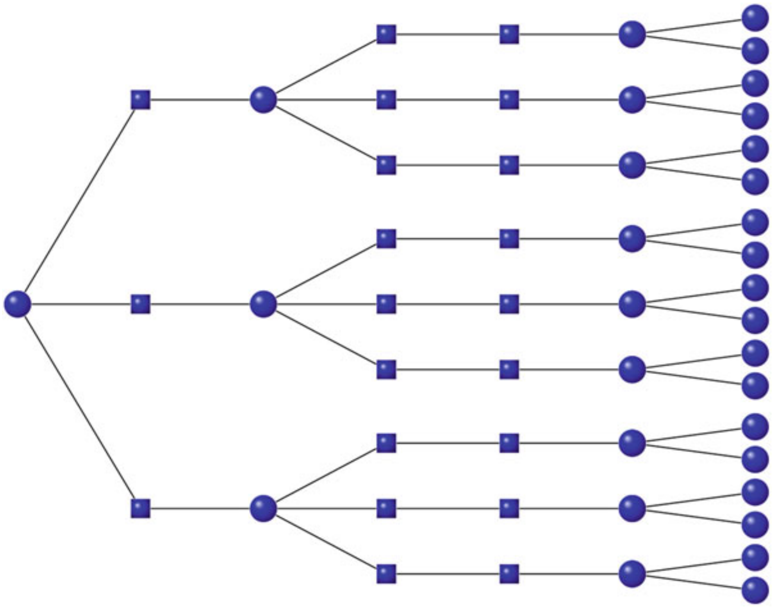


Fig. 13.1 An example of scenario tree with branching $3-(1)-3-(1)-(1)-2$

of the problem. As represented in Fig. 13.1, scenario branching will be allowed only at decision times (circle nodes) and not at intermediate times (squared nodes).

Accordingly we define the optimal investment policy, based on holding, buying and selling decisions along such tree. Let $X_n = \sum_i \sum_{h<n} x_{i,h,n}$ be the value of the investment portfolio in node n: i refers to the asset type and $h \in \mathcal{P}_n$ to a predecessor node of n where the investment originated. We consider four asset classes: I_1 for fixed-income and inflation-linked assets but corporates, I_2 for securitized and corporates, I_3 for equity and I_4 for real estate investments. The asset universe is defined by $I = I_1 \cup I_2 \cup I_3 \cup I_4$. In each class we have subsets of possible allocations for a total of 12 investment opportunities. The following decision variables are considered:

- $x^{+}_{i,n}$ — *investment* (buying decision) in node n, of asset i (a maturity T_i is considered for fixed-income assets);
- $x^{-}_{i,h,n}$ — *selling* decision in node n of asset i, that was bought in $h : (h \in \mathcal{P}^d_n) < n$; $t_n - t_h$ = holding period. In case of fixed-income assets, at the maturity T_i, the asset expires and it is assumed to generate a cash flow equal to the asset market value at maturity;
- $x^{exp}_{i,h,n}$ — *maturity* (expiry) at node n, of asset i (fixed-income asset) that was bought in h, $(t_n - t_h = T_i - t_h)$, $n \in \mathcal{N}_t$;
- $x_{i,h,n}$ — *holding* at node n, of asset i (with maturity T_i, for $i \in I_1$) that was bought in $h \in \mathcal{P}^d_n$;
- z_n — *cash account* in node n.

The asset value evolution in each decision node is captured by the return generated by previous decisions and current buying and selling decisions. We denote with $\rho_{i,n}$ the price returns at node n. Therefore, the main equations describing the PF portfolio evolution and inventory balance constraints of the ALM model are for $t \in \mathcal{T}, n \in \mathcal{N}_t$:

$$x_{i,n,n} = x_{i,n^+} \tag{13.1}$$

$$x_{i,h,n} = x_{i,h,n^-}(1+\rho_{i,n}) - x^-_{i,h,n} - x^{exp}_{i,h,n} \tag{13.2}$$

$$x^{exp}_{i,h,n} = x_{i,h,n^-}(1+\rho_{i,n}) \;\; h < n, \;\; n-h = T_i \tag{13.3}$$

$$x_{i,n} = \sum_h x_{i,h,n} \tag{13.4}$$

$$X_n = \sum_i x_{i,n} \tag{13.5}$$

Consider the random process $\rho_{i,n}$: for given holdings x_{i,h,n^-} of asset i in the parent node n^-, we derive the clean price return in node n. The intermediate argument $h \in \mathcal{P}_n$, representing the buying stage, is needed as discussed below to estimate the investment and losses over the holding period $t_n - t_h$. Those returns will determine the optimal investment allocation after paying the pensions. Unlike buying and selling decisions which are allowed only for $n \in \mathcal{N}^d_t$, bonds expiries and asset cashflows (interests and dividends) are recorded annually for each $n \in \mathcal{N}^{int}_t$. At the end of the time horizon sellings and purchases are not allowed, while fixed-income assets may expire and in this way generate cash-inflows: for $n \in \mathcal{N}^d_T, i \in I$ we have $x^+_{i,T} = x^-_{i,h,T} = 0$.

The portfolio manager is concerned with the issue of collecting sufficient resources to pay regularly pension obligations. In the long-run however she/he will focus on the PF funding status, namely the ratio of her asset portfolio to the liability. The common exposure to interest rate movements, historically, has been managed primarily by controlling the asset-liability (A-L) duration gap. We focus next on the ALM model liability-driven components and then analyse the funding issue.

13.3 Liabilities, Liquidity and A-L Duration Matching

The PF obligation in node $n \in \mathcal{N}_t$ for each $t \in \mathcal{T}$ is Λ_n. We assume known a given stream of net nominal pension payments $\tilde{L}_t$ over a long horizon $t \leq \hat{T}$ where $\hat{T}$ in insurance mathematics might be as large as several decades and such stream is just specified as an average pension obligation: the evaluation of the current liability implies first an inflation adjustment resulting into a tree process for pension cashflows and then at each node in the tree the discounting of future pension payments based on evolving term structures of interest rates. For the sake of simplicity and without introducing formally a stochastic interest rate model, at each decision node the current yield curve is assumed to be generated by fitting a term structure from the Nelson-Siegel (N-S) family (Nelson and Siegel, 1987) on the set of interest rate benchmarks adopted in the asset return model. A 1-year inflation rate π_n is generated as described below by difference between a nominal and real interest rates, resulting into inflation-adjusted random liabilities L_n for all $n \in \mathcal{N}_t$, where $t \leq \hat{T}$:

$$L_n = \tilde{L}_n \prod_{m \in \mathcal{P}_n} (1 + \pi_m). \tag{13.6}$$

At the 20-year horizon, the value of the net DBO B_n at leaf nodes $n \in \mathcal{N}_T$ defines an *end-effect* capturing the PF terminal funding conditions which will reflect the time T-forward net present value of all pension payments beyond T. Such evaluation will be performed relying on the yield curves on each leaf node.

We have, for $t \in \widehat{\mathcal{T}}$, $n \in \mathcal{N}_t$, indicating with t_n the reference time of node n:

$$\Lambda_n = \sum_{m \in \mathcal{C}_n} p_m \frac{L_m}{(1 + r_{t_n,t_m})^{(t_m - t_n)}}. \tag{13.7}$$

In (13.7) r_{t_n,t_m} denotes the interest rate quoted at t_n for deposits expiring in t_m, and $\forall m$, p_m define the conditional probabilities associated with the subtree originating from node n. At each stage equally probable transitions are assumed.

Term structure of interest rates Following Eq. (13.7), the discounting at each node relies on the fitting at time 0 and then along all scenarios, of a benchmark interest rate curve from the N-S family whose evolution is assumed to depend on a set of interest rate benchmarks spanning from 3-months to above 10-years durations. The N-S parametric model at time 0 for rates with maturity t reads:

$$r(0,t) = \beta_0 + (\beta_1 + \beta_2) \cdot \frac{\beta_3}{t} \cdot \left(1 - e^{-\frac{t}{\beta_3}}\right) - \beta_2 \cdot e^{-\frac{t}{\beta_3}} \tag{13.8}$$

where $\beta_0, \beta_1, \beta_2$ and β_3 are parameters to be fitted on market curves adopting a splining procedure. Yield curve calibration is performed relying on the N-S model on both a nominal and real interest rate curve: their difference will determine a term structure of (annual) inflation rates. The fitting is based on the evolution of the 3-month EURIBOR and Treasuries or Inflation-linked Treasuries with different maturity buckets (see below the asset return generating process). The implied 1-year inflation rate is then used to derive the inflation-adjusted pension process in Eq. (13.6).

The PF obligation in current terms will increase if inflation increases pushing up future pension payments and will decrease if interest rates increase, due to a decreasing discount factor. The two forces may offset each other as when interest rates remain equal in real terms.

Dedicated and duration-matching portfolios The PF manager at every node in the scenario tree seeks an asset portfolio generating the cash flows required by the stream (13.6). Furthermore, to neutralize undesirable effects that may be generated by volatile interest rates the duration of the asset portfolio should over time be as close as possible to the duration of the liability portfolio. We capture in the model those hedging conditions by introducing explicitly a liquidity gap and what we will refer to as *ALM* risk variable in the objective function.

To minimize cash imbalances in the pension service, every year those are funded through current asset income sources (interests and dividends) and bonds' redemptions. The income generated by bonds and equities over the decision horizon is identified by $\sum_{h<n,h\in\mathcal{P}_n^d} x_{i,h,n^-}\cdot\xi_{i,n}$, where $\xi_{i,n}$ denote random income coefficients along the scenario tree. Accordingly either a cash surplus will emerge, if cash inflows exceed pension payments, or a deficit will emerge: in either cases the resulting surplus/deficit will be compounded until the next decision stage and included in the cash balance constraint. We use L_n^Z to denote cash imbalances in intermediate nodes, $\forall\ t\in\mathcal{T},\ \ n\in\mathcal{N}_t^{int},\ \ h\in\mathcal{P}_n^d$ and $i\in I_1\cup I_2$:

$$L_n^Z=\sum_{h<n,h\in\mathcal{P}_n^d} x_{i,h,n^-}\cdot\xi_{i,n}+\sum_{h<n,\ n-h=T_i} x_{i,h,n}^{exp}-L_n \tag{13.9}$$

For model accuracy we recall that, given the peculiar tree structure in Fig. 13.1, to preserve non-anticipativity, neither buying nor selling decisions are possible in intermediate stages. Equation (13.9) jointly with the cash balance constraint at decision stages clarify how a *dedicated* portfolio approach is captured in the model formulation. At time 0, the cash account is defined by the cash inflows due to selling decisions $x_{i,h,n}^-$ and outflows due to new investments $x_{i,n}^+$. Given an initial cash balance z_{0^-}, we have $z_0 = z_{0^-}+\sum_i x_{i,0^-,0}^- - \sum_i x_{i,0}^+$. Afterwards at each decision time the cash account will depend on pension payments L_n, on compounded cash imbalances $L_{h,n}^z$, on cash inflows generated by the investment portfolio and on potential *sponsor*'s contributions Φ^u. The latter, as clarified below, are regarded as last-resort, costly liquidity injections to fund temporary liquidity deficits.

$$\begin{aligned} z_n = z_{n^-}&(1+\zeta_n)+\Phi_n^u+ \\ &-L_n+\sum_{n^{d-}<h<n} L_{h,n}^z(1+\zeta_{h,n})+ \\ &+\sum_i\sum_{h<n,h\in\widehat{\mathcal{P}}} x_{i,h,n}^- -\sum_i x_{i,n}^+ \\ &+\sum_i\sum_{h<n,\ n-h=T_i} x_{i,h,n}^{exp}+\sum_{i\in I\setminus I_1,I_2}\ \sum_{n^{d-}<m<n,\ n-m\geq T_i} x_{i,m,n}^{exp}(1+\zeta_{h,n}) \\ &+\Pi_n^{1,INV} \end{aligned} \tag{13.10}$$

The quantity $\Pi_n^{1,INV}$ represents interest and dividend income and it is computed in the model as:

$$\Pi_n^{1,INV} = \sum_{i\in I}\sum_{h<n,h\in\mathcal{P}_n^d}\sum_{n^{d-}<m<n,m\in\mathcal{P}_n^{int}}\left(x_{i,n^{d-}}\cdot(1+\rho_{i,m})-x_{i,h,m}^{exp}\right)\cdot\xi_{i,n}+ + \sum_{i\in I}\sum_{h<n,h\in\mathcal{P}_n^d} x_{i,h,n^{d-}}\cdot\xi_{i,n}. \tag{13.11}$$

In Eq. (13.11) we compute at decision nodes $\mathcal{P}_n^d$ the income generated by holdings at the parent decision node $x_{i,h,n^{d-}}$ and that generated by holdings that didn't expire and were held during preceding intermediate stages.

Along each scenario, consistent with the assumed tree structure, cash surpluses and deficits will be passed forward to the following stage together with the accrual interest. Very low positive interest rates ζ_n^+ and penalty negative interest rates ζ_n^- will force the investment manager to minimize cash holdings over time. The cash surplus at the end of the horizon is part of company terminal wealth. In the adopted problem instance we do now allow cash account deficits over the horizon. The PF sponsor will always intervene in case of shortages.

Further to holding a portfolio generating sufficient cash to face current liabilities, th PF manager wishes to minimize exposure to interest rate fluctuations by maintaining a sufficiently narrow A-L duration gap. The *ALM Risk* in node n as generated by an A-L duration gap in the parent node is denoted by K_n^1 to indicate that it is the primary variable contributing to the definition of the PF interest rate risk exposure:

$$K_n^1{=}dr^+\cdot\left(t_j-t_{j-1}\right)\cdot\left(\Delta_n^x-\Delta_n^\Lambda\right)_+ - dr^-\cdot\left(t_j-t_{j-1}\right)\cdot\left(\Delta_n^\Lambda-\Delta_n^x\right)_+ \tag{13.12}$$

where $dr^{+/-}$ are positive and negative interest rate changes and $\Delta_n^x-\Delta_n^\Lambda$ defines the difference between asset and liability durations. Only the linear first-order interest rate sensitivity on assets and liabilities is considered here. The ALM, or interest rate risk will increase if an exogenous positive interest rate shock occurs when asset duration exceeds liability duration, while it will decrease in the opposite case.

Summarizing, we handle the two risk sources – liquidity and interest rate risk – by introducing in the objective function the sum of *liquidity gap* and ALM risk:

$$\Psi_n{=}\Omega_n+K_n^1+\Psi_{n^-} \tag{13.13}$$

where $\Psi_{t_0}=0$, K_n^1 is the ALM risk, while the liquidity gap Ω_n is defined as:

$$\Omega_n = L_n - \Pi_n^{1,INV} - \sum_{h,t_{n^{d-}}<h<t_n,} L_h^Z(1+\zeta_h) - \sum_{h<n,\ n-h=T_i} x_{i,h,n}^{exp}. \tag{13.14}$$

The PF manager aims at minimizing Ψ_n over the planning horizon. In addition she is expected to control the portfolio market risk exposure: K_n^1 reflects the interest rate exposure, while K_n^m will reflect the exposure of the asset portfolio to extreme market losses. By extending the Sharpe ratio concept to tail risk, the PF manager is assumed in the long-run to seek a strategy such that the portfolio return per unit K_n^m risk is maximised.

13.4 Risk Capital and Risk-Adjusted Performance

By dynamically rebalancing the portfolio the PF manager will continuously revise the risk exposure and seek a higher risk-adjusted return. We assume a rather simple, highly conservative, constant correlation model to clarify how such exposure is kept under control over the decision horizon. Indeed it can be shown that under assumptions of perfect positive correlation between assets' risk factors, following Consigli and Moriggia (2014), not only are we considering a worst-case scenario from a financial viewpoint but the resulting MSP will be linear:

$$K_n^f = K_n^1 + K_n^m + K_{n-}^f \tag{13.15}$$

where $K_{t_0}^f = 0$, K_n^1 is the ALM risk, while K_n^m evaluates the linear market risk exposure as:

$$K_n^m = \sum_i \sum_{h<n, h \in \mathcal{P}_n} x_{i,h,n} \cdot k_i \cdot (t_n - t_{n-}). \tag{13.16}$$

The asset-specific coefficients k_i describe the extreme returns-at-risk associated with holdings in asset i over a unit period (say 1 year). These risk coefficients, focusing on tail returns, may be determined either through an appropriate statistical approach or by introducing regulatory-based risk charges. The latter approach is taken here with k_i expressing the 99% returns at risk associated with different asset classes. Such approximate simple risk estimation is assumed to be roughly consistent with the return generating processes for price and income returns, needed by the model instantiation, as shown next. The aggregate tail risk exposure of the pension fund is defined by the sum of the financial and the actuarial risk capital $K_n = K_n^f + K_n^l$, where K^l denotes an estimate of possible extreme losses generated by liability dynamics, due to unexpected longevity records or unprecedented lump-sum pension requests. This is assumed to be defined as a (small, say 5%) percentage of the PF obligation: $K_n^l = \phi^l \cdot \Lambda_n$. In the definition of the optimal strategy the PF manager is assumed to seek a return per unit tail risk target based on:

$$Z_n = \frac{\Pi_n^{cum}}{K_n^f}. \tag{13.17}$$

The portfolio's return per unit tail risk is defined as a ratio between the total profit generated by the investment (including the variation of unexpected gain and losses) and the investment risk exposure. To avoid handling a nonconvex variable within the optimization problem, rather than maximizing the ratio we maximize the difference $Z_n = \Pi_n^{cum} - K_n^f$. The investment risk capital K_n^f at each stage is computed assuming perfectly positively correlated asset classes and then deriving from the optimal portfolio generated by the solution the dynamics of the risk capital. We define the total portfolio return by cumulating period price returns, upon selling decisions, and unrealized gain and losses from the holding portfolio.

$$\Pi_n^{cum} = \Pi_{n-}^{cum} + \Pi_n^{INV} + [UGL_n - UGL_{0-}] \tag{13.18}$$

where $\Pi_n^{INV} = \Pi_n^{1,INV} + G_n$ is the sum of portfolio income and trading profits. The unrealized gain and losses UGL are defined by $UGL_n = \sum_i \sum_{h<n, h \in \mathcal{P}_n} x_{i,h,n} \cdot \chi_{i,h,n}$ with $\chi_{i,h,n}$ to denote an unrealized gain-loss coefficient in node n per unit investment in asset i in node h. Realized gain and losses G_n are instead determined by clean price variations upon selling and upon assets redemptions:

$$G_n = \sum_i \sum_{h<n, h \in \mathcal{P}_n^d} \left(x_{i,h,n}^- + x_{i,h,n}^{exp} \right) \cdot g_{i,h,n} \tag{13.19}$$

The coefficients $g_{i,h,n}$ quantify unit gains-losses upon sellings assets i in node n which was bought in node h. Both realized and unrealized gains depend on the outcome of a return process for given rebalancing decision: at the beginning of each stage an asset allocation must be selected whose outcome at the end of the stage will depend on realized returns. We summarize the adopted Gaussian return model before discussing the objective function and the dynamic risk-reward tradeoff considered in this study.

Return generating processes Assets' random return coefficients are derived by simulation along the scenario tree. From a methodological viewpoint we apply first a simple Monte Carlo method to determine the risk factors evolution over a simulation horizon, then assume a given tree process and generate the returns' nodal realizations by sampling consistently with the conditional tree structure. We are not applying any specific sampling method (Consigli et al., 2012a; Dempster et al., 2011; Dupačová et al., 2001) but just focusing on a simple scenario generation approach to evaluate the ALM model and the PF induced funding policy. For each node n in the tree $\rho_{i,n}$ is the price return of asset i in node n, $\xi_{i,n}$ is the income return and ζ_n indicates the cash account return in node n. Each return type is assumed to be associated with a specific benchmark $V_{i,n}^j$ where $j = 1, 2, 3$ to distinguish price from income and money market returns respectively, while i refers to the asset class. We have $v_{i,n}^j := \frac{V_{i,n}^j - V_{i,n-}^j}{V_{i,n-}^j}$ and

$$v_n^j = \mu^j \Delta t_{n-} + \Sigma^j \Delta W_{t_{n-}}. \tag{13.20}$$

In (13.20) v_n^j is a return vector whose dimension coincides with the asset universe cardinality, with mean μ^j and variance-covariance matrix Σ^j. Δt_{n-} denotes the time increment between node $n-$ and n while $\Delta W_{t_{n-}}$ denotes a white noise random vector. We have $v^j \sim N(\mu^j, \Sigma^j)$. The set of coefficients actually adopted in the case study can be found in the appendix. The following investment opportunities are considered: for $i = 0$, $j = 3$: the EURIBOR 3 months; for $j = 1, 2$, $i = 1, 2, 3, 4, 5$: the Treasury indices for maturity buckets 1–3 years, 3–5 years, 5–7 years, 7–10 years and 10+ years, respectively; for $i = 6$: the Securitised Bond index; for $i = 7$: the Corporate Investment Grade index; for $i = 8$: the Corporate High Yield index; for $i = 9$: the Real Estate Indirect index;for $i = 10$: the Public equity index; for $i = 11, 12$: the Treasury Inflation Protected Securities (TIPS) indices for maturity buckets 3–5 years and 10+ years, respectively. Treasury indices and TIPS are used respectively to infer annual interest rates for the nominal and real yield curves of the N-S model in Eq. (13.8).

We consider 4 asset classes: I_1 includes all fixed-income and inflation linked assets, I_2 the securitised asset and the corporate, I_3 the public equity investment and I_4 the real estate.

Each asset i is characterized by a return process and in addition by a tail risk coefficient k_i. From above, leaving aside the 3-month interest rate, all other assets have associated both a clean price return process $\rho_{i,n}$ and an income process $\xi_{i,n}$. We indicate with $\rho_i \sim N(\mu^i, \sigma^i)$ the return i marginal distribution, in what follows we assume that the regulatory-based risk coefficient k_i is determined from Solvency II regulatory estimates by an updating based on historical estimates. As such they will just provide a maybe conservative estimate of the 99% return-at-risk associated with ρ_i. The following tail risk coefficients are considered in this study: $k_i = \{2.5\%, 4\%, 5\%, 4\%, 7\%, 20\%, 39\%, 25\%, 4\%, 3\%\}$ to identify potential 1-year price return tail losses associated with Treasuries (1–3, 3–5, above 5 years), Securitized bonds, Investment- and Speculative-grade corporates, equity, real estate and short- and long-term TIPS, respectively.

13.5 Funding Conditions and ALM Optimization

The ALM problem is formulated as a MSP recourse problem (Birge and Louveaux, 1997; Consigli and Dempster, 1998) over six stages. The optimal root node decision is taken at time $t_0 = 0$ and, from a practical viewpoint, represents the key implementable decision. Recourse decisions occur at t_1, t_2, t_3, t_4 and t_5 which represent respectively 1st, 2nd, 3rd, 5th and 10th year. At t_6, the $20 - year$ horizon, no buying or selling decisions are allowed but assets expiries and income leverages from the holding portfolio.

The PF manager has a primary interest to run the pension plan smoothly, match all the funds' liabilities and allow retired employees to keep a sufficiently good living standard during retirement. Of major concern to PF managers is any condition of liquidity shortage that may result into extrafunding from the fund's sponsor. In practice this is regarded as a last-resort highly undesirable option. The ALM problem objective function combines those elements to yield jointly a good hedging strategy and the generation of risk-adjusted extra returns. Through the decision horizon the portfolio manager will monitor carefully the net DBO $B_n = \Lambda_n - X_n$ defined by the difference between the present value of the accrued pension obligation Λ_n and the fair value of plan assets X_n in all nodes n of the scenario tree.

The net DBO is directly related to the FR $\Gamma_n = \frac{X_n}{\Lambda_n}$. These two quantities represent the key drivers of the PF long-term policy. A PF underfunding condition is expressed by a FR below 1, above 1 for an overfunding condition. Regulators pay now-a-day increasing attention to avoid any undesirable risk-taking by PF managers and specifically structural underfunding conditions. In the case study we show that the adopted ALM model formulation proves effective to drive a PF from an under to an overfunding condition: it is of interest to study how such outcome is associated with a given risk control strategy.

The objective function is based on a convex combination of the liquidity gap plus the ALM risk ($\Psi_n = Y_{1,n}$), the return on plan assets per unit tail risk ($Z_n = Y_{2,n}$), the plan sponsors' unexpected contributions ($\Phi_n^u = Y_{3,n}$), and the net defined benefit obligation (DBO) ($B_n = Y_{4,n}$):

$$\min_{x \in X} \left[\sum_{j=1,2,3,4} \lambda_j \mathbb{E} \left(\widetilde{Y}_j - Y_{j,n} | F_j \right) \right] \tag{13.21}$$

with $\sum_{j=1,2,3,4} \lambda_j = 1$.

In (13.21) further to Y_j, $\widetilde{Y}_j$ identify the exogenous j targets, and $\mathbb{E}(Y_{j,n}|\mathcal{F}_j)$ denotes conditional expectation with respect to information sets $\mathcal{F}_j$ associated with target $\widetilde{Y}_j$.

The optimal decision sequence x^* of the problem is defined over the decision stages $x \in X, x := \{x_t\}_{t \in T_d}$ only and will not include intermediate nodes, used instead to manage annual pension payments. By including the above set of targets, the PF management seeks a good trade-off between a dedication and duration-matching strategy, as reflected in $Y_{1,n}$, a goal risk-adjusted return $Y_{2,n}$, a long-term funding surplus $Y_{4,n}$ and avoid or minimize sponsor's unexpected contributions $Y_{3,n}$.

The ALM model develops from a canonical liability-driven investment approach to incorporate a risk-adjusted return variable and an extended set of constraints. Inflation and interest rate risk play a central role in determining liability dynamics. Interest rates and market risk factors affect instead the asset portfolio.

Overall the ALM model risk-reward trade-off considers the following risk sources, whose control is key to the maintenance of a sufficient funding status.

- Liquidity risk: annually the asset portfolio generates cash-inflows that jointly with incoming pension contributions are used to pay current pensions. In case of positive asset-to-passive PF members ratio, contributions may very well be sufficient to fund current pensions, but otherwise a liquidity gap will emerge, defined by the difference between net pension payments and portfolio income.
- Inflation risk: DB occupational PFs' liabilities are typically (even if not always) inflation-adjusted and this is a primary risk source affecting PF liabilities, jointly with interest rate risk.
- Interest rate risk: PF liabilities carry a payment term structure. The DBO joint dependence on interest rates and inflation leads to exposure to so-called real duration mismatching. Accordingly the PF manager will hedge such exposure by compensating liability real duration on the asset side. Real duration may be considered as first-order sensitivity of assets and liability values to changes of real interest rates.
- Market (tail) risk: Increasing volatility in financial markets facilitated the penetration of advanced risk management practice in the historically traditional PF industry. Specifically in the case of occupational PFs within financial and insurance conglomerates, underfunding risk has an impact on the sponsor's financial equilibrium. The PF manager, consistently with Solvency II regulations, seeks also an optimal control of extreme market risks: this is based in the ALM model on the introduction of a set of risk coefficients under an assumption of perfect positive correlation between risk factors.
- Longevity risk: In case of underestimated passive members survival probabilities, upon conversion at retirement time of the pension credit into an annuity may in the long-run weaken the PF solvency resulting into longer than expected pension payments. We assume that such risk is not affecting the PF and already accounted for in the liability estimation.

A negative PF scenario may result from joint negatively correlated shocks to assets and liabilities resulting into a sudden increase of liabilities matched by a decrease of asset values. In presence of poor duration matching and excessive exposure to market risk on the asset side, a shock in the equity market is often associated with easing conditions in the money market and increasing liability values.

A poor reaction to market risk is sometimes induced by portfolio policy constraints, including: lower and upper bounds on asset positions, a turnover constraint limiting portfolio revisions whatever the market condition, and liquidity constraints. These are considered in the ALM model instance and here summarized.

For $t \in \mathcal{T}$, $n \in \mathcal{N}_t^d$, $h \in \mathcal{P}_n^d$, and $i \in I$ we have:

$$l_i X_n \leq x_{i,n} \leq u_i X_n \tag{13.22}$$

where l_i and u_i define the percentage lower and upper bounds on holdings in asset i with respect to the current portfolio position.

For $t \in \mathcal{T}$, $n \in \mathcal{N}_t^d$, $h \in \mathcal{P}_n^d$, and $i \in I$, a maximum turnover constraint:

$$\sum_{i \in I} \sum_{h<n} x_{i,h,n}^{-} + \sum_{i \in I} x_{i,n}^{+} \leq \vartheta \left[\sum_{i \in I} x_{i,n-} - \cdot (1 + \rho_{i,n}) \right] \tag{13.23}$$

where ϑ is a predefined turnover coefficient. In our analysis the turnover constraint implies that the volume of the investments and disinvestments, at each node n, must be less than or equal to a certain percentage of the portfolio value before any rebalancing.

Finally a liquidity bound:

$$\sum_{i \in I} \left(X_{i,n} \cdot \widehat{\ell}_i \right) \geq X_n \cdot \widehat{\ell} \tag{13.24}$$

where $\widehat{\ell}$ is a 50% input coefficient.

After considering the set of assets' policy constraints, we summarize the key elements of the ALM model and their rationale from an economic viewpoint:

1. A distinctive modeling feature is represented by the introduction of intermediate nodes between subsequent decision nodes in the scenario tree, where however, the non-anticipativity of investment decisions is preserved and we may accommodate annual frequency of pension payments and liability evaluation.
2. The model integrates dedication and AL duration matching with the introduction of a risk-adjusted return measure and a net DBO target: both determine a strong incentive to the PF manager to maximize compounded portfolio returns and at the same time minimize the exposure to financial risk.
3. The definition of intermediate stages allows a more accurate mapping of liquidity conditions and by doing so a sufficiently safe estimation of the sponsor's extraordinary contributions, if any.
4. A perfect positive correlation between assets' returns is assumed in the estimation of the investment portfolio risk exposure, leading to a conservative risk capital estimation (thus forcing the statistical estimates used to generate assets' returns). The investment universe is agreed with experts from the insurance world and includes Treasuries, TIPS, corporates, equity and real estate.
5. Fixed income benchmarks are treated in the model as carrying a maturity and generating price as well as income returns. Real estate as well as TIPS investments, sometimes referred to as *hard* assets play historically a central role in the PF's inflation hedging strategies.

13.6 Case Study: A 20 Year Pension Fund ALM Problem

In this final section we analyse a case problem focusing directly on the conditions for a DB pension fund to recover a positive funding surplus from an initial FR severely below 1. To do this we solve the optimization problem under operational assumptions and after ranking scenarios in terms of net DBO evolution we consider first the FR evolution across all scenarios, then a strategy employed to achieve a FR closed to or greater than 1 specifically under a worst-case-scenario.

Initial conditions include a discounted value of pension payments (from year 1 to the far future) Λ_0 estimated at EUR 115 million (mln), a fair value of plan assets A_0 equal to EUR 100 mln and expected net pension payments in the first year of EUR 35 mln. Accordingly at time 0 the fund has a FR around 87%. The initial portfolio X_{0-} is assumed to be well-diversified across the investment universe. The PF liquidity condition is also delicate since an annual shortage of approximately EUR 31 mln can be anticipated assuming a portfolio income return of 4% in the first year. Under this condition the PF manager wishes to recover over the forthcoming 20 years a funding surplus. We assume a close fund in which members can only decrease over time. As pensions are paid, the PF liability will decrease: inflation scenarios and longevity will however induce a relatively slow decreasing path. The case problem develops from an input expected stream of future annual pension payments, provided directly by the actuarial division of an occupational PF, by generating inflation-adjusted pension flows and then at each node by discounting future cash-flows using nodal specific term structure of interest rates. All data in this case-study are modified and rescaled for confidentiality reasons but agreed with industry colleagues and representative of real-world business conditions.

The optimal investment policy is constrained by a set of upper and lower bounds defined with respect to the nodal portfolio values. No short positions are allowed over the planning horizon. Treasury and TIPS are unconstrained, as well as corporate fixed income investments. The equity holdings as well as real estate holdings cannot exceed 30% of the portfolio and a 30% turnover constraint with a minimum 30% liquidity bound are assumed.

The following (decision) tree is considered in Table 13.1. The current implementable decision, corresponding to the root node, is set to January 1, 2015.

Table 13.1 Pension fund case study time and space discretization

Decision stage	1	2	3	4	5	6	7
Stage distribution (years)	0	1	2	3	5	10	20
Branching degree	10	4	2	2	2	2	
Scenarios		10	40	80	160	320	640

Between two decision stages, maybe distant in time, we consider annual intermediate times at which income returns and expiring assets are used to pay annual pension payments. Any liquidity surplus or shortage at intermediate stages is compounded to the next decision stage.

The objective function (13.21) of the PF ALM problem considers a trade-off between different goals through the coefficients $\lambda_1 = 0.2$, $\lambda_2 = 0.2$ and $\lambda_3 = 0.4$ $\lambda_4 = 0.2$. We assume that the pension fund's management will revise its strategy so as to minimize the shortfall from given targets of (1) liquidity gap plus ALM risk, (2) return per unit tail risk, (3) plan sponsors' unexpected contributions and (4) net DB. The following targets are assumed: EUR 7 mln for (1), a 2% per annum for (2), 0 for (3) and a EUR 4 mln of funding surplus.

Rebalancing decisions can be taken at decision stages from time 0 up to the beginning of the last stage; no decisions are allowed at the end of planning horizon. The optimal decision sequence does not include intermediate decisions. Nevertheless, pension payments and assets' returns will affect the revenues and asset-liability streams.

The results are generated through a set of modules combining Matlab R2011b as the main development tool, GAMS 23.2 as the model generator and solution algorithms interface and Excel 2010 as the input and output data collector running under a Windows 10 operating system with 8 GB of RAM and a dual core with four logical processors. The MSP is generated through GAMS and the CPLEX dual simplex algorithm is used to solve it. With 640 scenarios this is a very large-scale optimization problem with 1.344.995 rows, 1.464.345 columns and 7.659.404 nonzero coefficients, which is generated in roughly 16 min and solved in 5:29 (minutes:seconds) CPU time.

Consider along a specific representative scenario, the evolution of the net DBO and the FR in Fig. 13.2. The net DBO over the first years is positive reflecting a negative PF condition and accordingly the FR is deeply below 1, then between years 15 and 20 the portfolio strategy leads to a surplus at the horizon with a negative net DBO and a FR above 1.

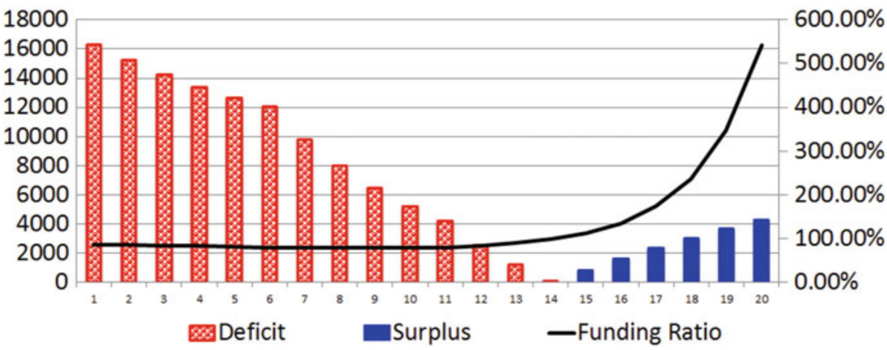


Fig. 13.2 Funding ratio (right Y axis) and net DBO (left Y axis, thousand EUR) mean scenario

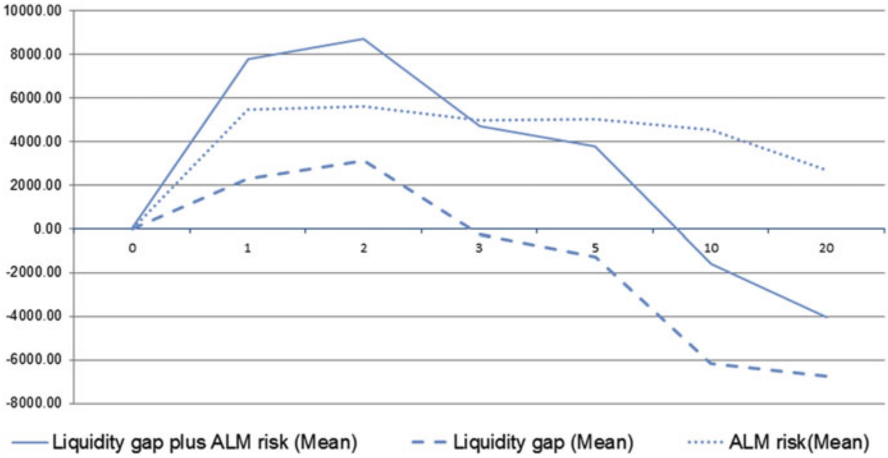


Fig. 13.3 Liquidity gap and ALM risk (thousand EUR)

A condition of underfunding, FR below 1, due to assets versus liabilities cashflows, may be associated with a liquidity surplus in the short term. Here however along the same representative scenario a condition of liquidity shortage emerges with a positive exposure to interest rate risk. Indeed the liquidity gap is positive over the first 3 years but then it starts decreasing until a liquidity surplus emerges. The exposure to interest rate risk also remains low as shown in Fig. 13.3 to witness the effectiveness of the joint risk control pursed by the optimal strategy over the planning horizon.

We see below that despite such negative initial conditions the PF will recover a robust funding condition: such outcome is primarily induced by decreasing liabilities and an effective disinvestment strategy leading to increasing risk-adjusted investment profit and terminal FR above 1 across all scenarios, but at the cost of a significant reduction of the asset portfolio. Indeed even in the worst case scenario, a decreasing liability scenario, due to outgoing pension fund members and relevant lump-sum payments in the first stages, is funded through income returns and asset sellings without the need of sponsor's intervention.

In this case study, we concentrate primarily on the FR evolution over time and across all the scenarios. The plots in Fig. 13.2 refer as indicated to an average scenario: this is identified, after the problem's solution, by ranking scenarios with respect to the evolution of the net DBO over the entire planning horizon. Out of 640 scenarios, the mean case of all is identified by considering the scenario that at each stage remains closer in the Euclidean norm to the mean.

The chance to recover a positive funding condition depends, for given exogenous liability scenarios, entirely on the investment strategy. Portfolio revisions take into account the need to preserve a sufficient liquidity buffer and minimize sponsor's interventions. We show that indeed these goals are attained over the 20 years in all scenarios.

13.6.1 Evolution of Funding Conditions

We present in Fig. 13.4 the evolution of the net DBO across time and scenarios: at the end of the first stage we show net DBO values in each of 10 nodes in decreasing value from left to right. At the end of the second the associated descending 40 nodes and so forth until the horizon. Red colour implies positive net DBO, thus liabilities exceeding asset values, blue colour is instead associated with a negative net DBO with the asset portfolio now exceeding the corresponding nodal DBO estimates.

At the end of the 10th year, already, only a limited but large portion of the scenarios carry a negative though rather high net DBO: such evidence is of extreme interest to the PF manager because under that condition all regulatory and long-term financial and risk constraints are satisfied. At the 20 year horizon across all 640 scenarios the net DBO is negative: in each leaf node the terminal value of the asset portfolio exceeds the discounted value of future pension payments. Net DBO leaf nodal values thus depend jointly on the terminal investment portfolio and for given future inflation scenarios on the prevailing yield curve. The EUR 4 mln funding surplus is achieved in all scenarios as shown in Stage 6.

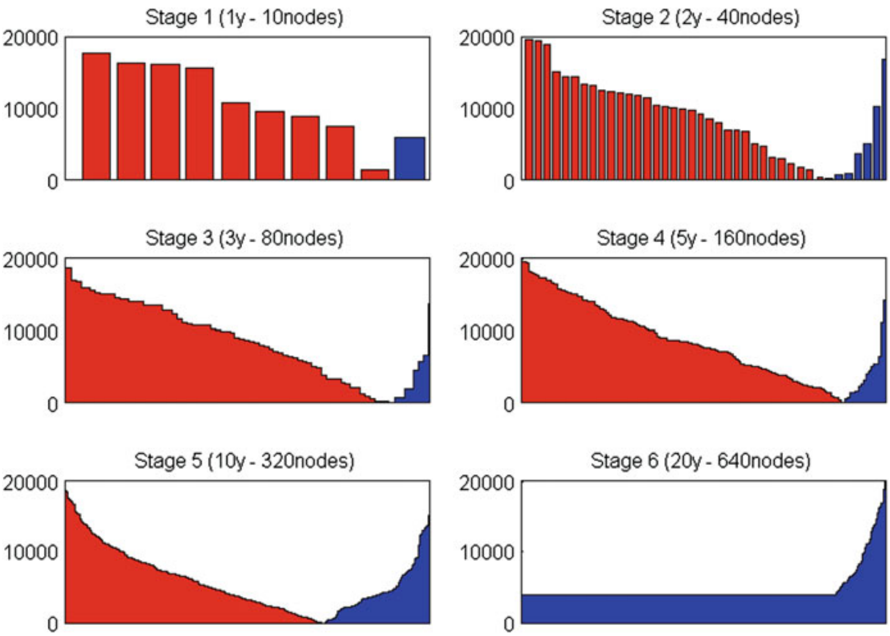


Fig. 13.4 Net DBO (Y axis, thousand EUR) at stage nodes (X axis)

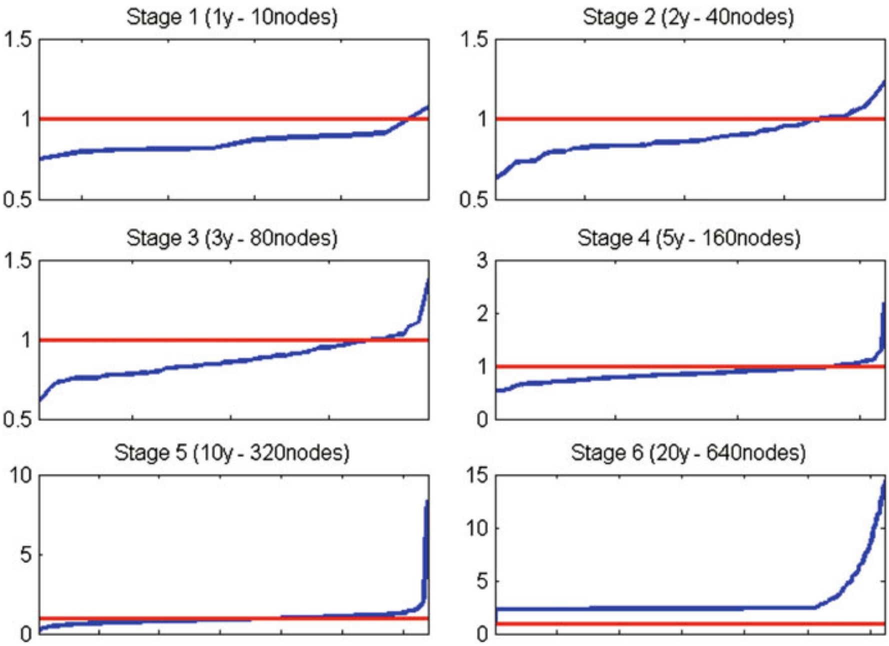


Fig. 13.5 Funding ratio (Y axis) at stage nodes (X axis)

The evolution of the FR can be analyzed under the same graphical structure. In Fig. 13.5 we show the corresponding stage-by-stage FR nodal values: the constant red line in each plot indicates an equilibrium condition: from the first to the last stage across all scenarios the FR moves from below to above 1.

The evidence reported in Figs. 13.4 and 13.5 highlights the achievement of a funding surplus across all scenarios. The ALM solution refers to ex-ante information and it supports the claim that under the given initial conditions and input scenarios, the problem solution leads to full recovery of a funding surplus consistently with the PF managerial goals. Such surplus is achieved satisfying risk capital and policy constraints.

We provide in Fig. 13.6 a final set of evidence on targets' achievements over the 20 years. We indicate in light colour the percentage of scenarios at each relevant stage where the goals were achieved or even overachieved and in dark those scenarios for which that target wasn't achieved. At the 1 year horizon a positive liquidity gap and ALM risk persist in all scenarios and from the given initial shortage condition the PF recovers a good liquidity status after 10 years. Nevertheless, as shown on the top right plot, all liquidity deficits are funded avoiding any sponsor's

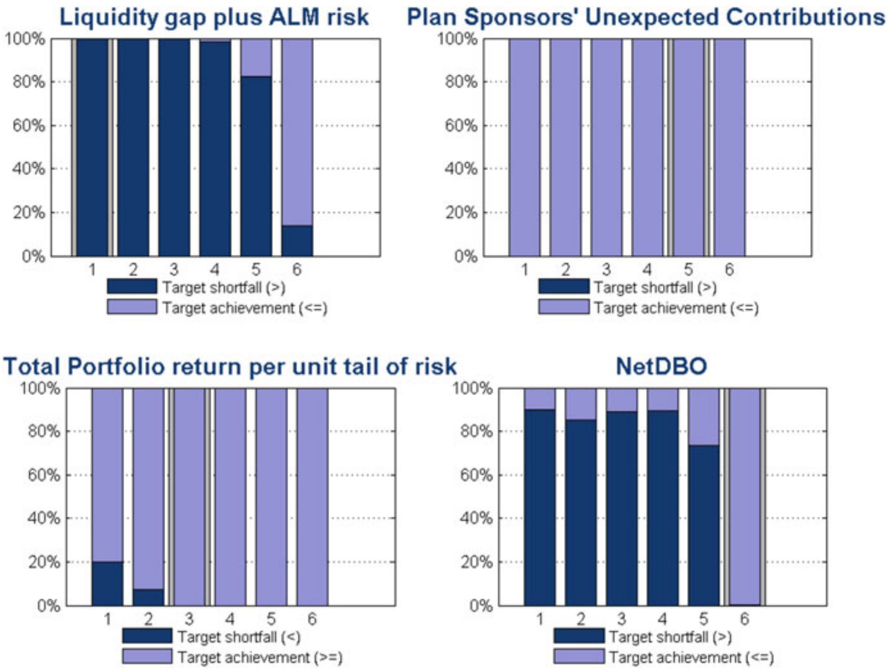
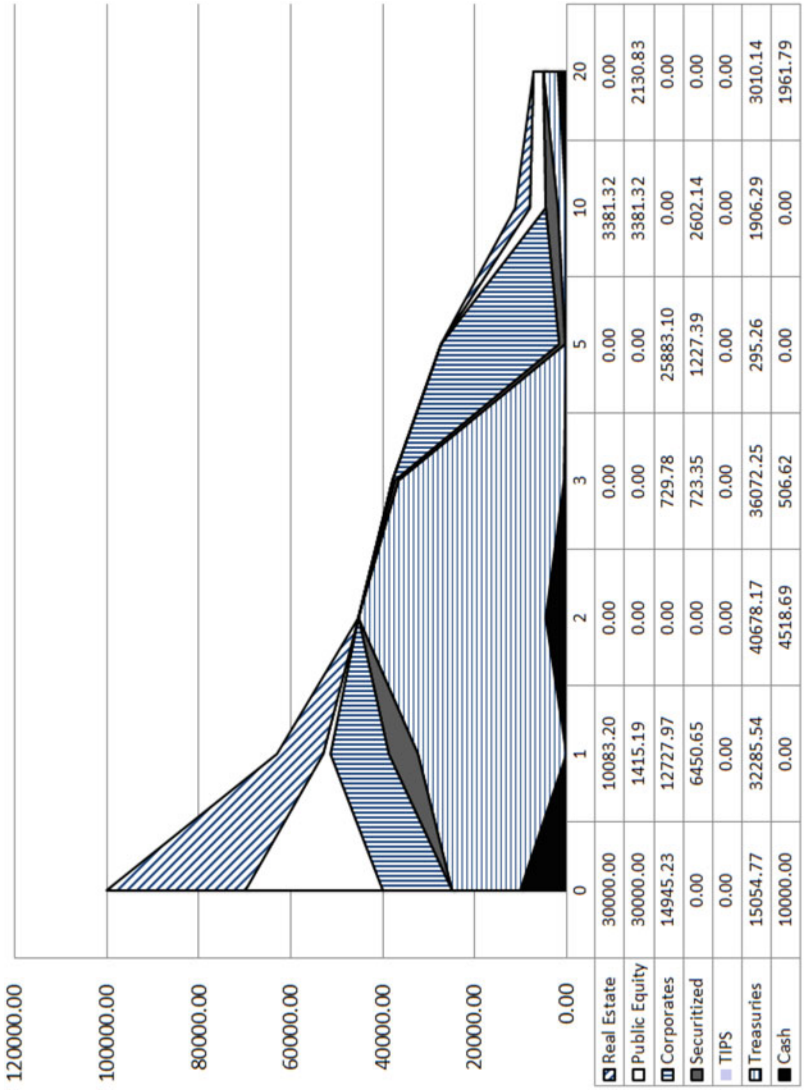


Fig. 13.6 Targets achievement

contributions which remain equal to 0 throughout the decision horizon and through assets sellings (see evidence in Fig. 13.7). The 2% risk-adjusted return target is achieved in all scenarios starting from the 3-year horizon. Finally (bottom right plot), as already discussed, progressively over time the PF recovers a long-term funding equilibrium condition.

Two relevant issues arise: which key elements drive the PF into such long-term equilibrium condition? And to which extent is such outcome due to the proposed modeling and optimization approach? Relatively in this latter case, to other approaches currently used in the industry.

As for the first question, we analyze in Sect. 13.6.2 which portfolio policy, under the worst possible DB scenario, leads to a full recovery of funding conditions over the two decades. We argue that the introduced targets' combination and a large investment universe lead, within a dynamic formulation, to an efficient dynamic portfolio diversification together with a very effective risk control. Extensive computational experiments, furthermore, have shown that such outcome may have been difficult to achieve without including in the objective function the net DBO and the risk-adjusted return goals. Indeed, the collected evidence supports the claim that a duration-based short-term hedging goal is consistent with a medium- to long-term more aggressive investment policy. A tight liquidity management of the PF thanks to the intermediate stages, also played across most experiments a relevant role.



	0	1	2	3	5	10	20
Real Estate	30000.00	10083.20	0.00	0.00	0.00	3381.32	0.00
Public Equity	30000.00	1415.19	0.00	0.00	0.00	3381.32	2130.83
Corporates	14945.23	12727.97	0.00	729.78	25883.10	0.00	0.00
Securitized	0.00	6450.65	0.00	723.35	1227.39	2602.14	0.00
TIPS	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Treasuries	15054.77	32285.54	40678.17	36072.25	295.26	1906.29	3010.14
Cash	10000.00	0.00	4518.69	506.62	0.00	0.00	1961.79

Fig. 13.7 Optimal dynamic allocation worst scenario (Y axis, thousand EUR)

As for the second question, the benefits carried by a dynamic optimization approach are undoubtable in presence of long-term liability streams. Widely in the European and U.S. PF's industry, and even earlier in the U.K. (Franzen et al., 2007) investment managers started moving over the last two decades towards the adoption of dynamic modelling approaches and long-term scenario analysis. Either based on Monte Carlo methods often matched by policy rules optimization (Mulvey et al., 2008) or on stochastic optimization (Dert, 1998; Consigli and Dempster, 1998) techniques. From a recent survey (Senders, 2010) a majority of European PF managers is now-a-days concerned with an accurate and statistically sound liability projection based on internal risk models and increasingly with an effective estimation of both asset and liability risk exposure estimation. The approach here reported, however, is the first application of a multi-criteria objective function over such long-term horizon with annual liability cash-flow matching.

13.6.2 Worst Case Scenario Analysis

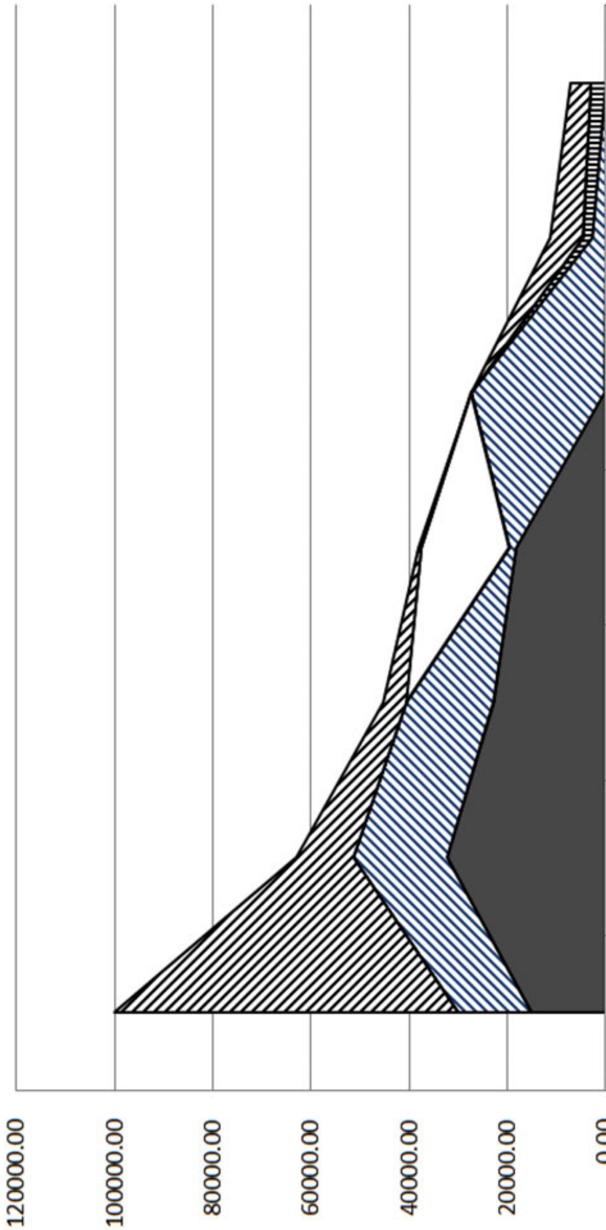
We move from a mainly quantitative analysis to a more qualitative one by analyzing the adopted optimal investment strategy along a specific, particularly negative, FR scenario.

In Fig. 13.2 we have shown the evolution of liquidity and funding conditions along a specific average scenario: that represented a mean-case-scenario from the perspective of the net DBO dynamics after solution. Indeed under that scenario the FR was around 80% for many years then decreased to less than 70%, an already warning FR, and finally it recovered to almost 120% after 15 years and kept increasing until the horizon.

Not only under the average scenario but along all scenarios the asset portfolio value decreases from the initial 100 mln value over the planning horizon. In the following we analyze the optimal portfolio strategy under worst-case FR conditions, where however the terminal funding surplus is also achieved.

The optimal H&N solution is well-diversified at time 0 with, however, equity and real estate at their upper bounds. As shown in Fig. 13.7 over the first 3–4 years the optimal strategy concentrates on treasuries and fixed-income investments and then increasingly on real estate and equity beyond the 5 year horizon. Such strategy is consistent with the achievement of a minimal shortfall with respect to liquidity gap and ALM risk in the short term and then increasingly in search of price gains and returns. Real assets seem to be contingently invested depending on the evolution of the risk-bearing capacity of the PF.

In Fig. 13.8 we see that the short-term duration matching goal and the medium term risk-adjusted performance goal rely primarily on a portfolio including assets with no expiry (equity and real estate) plus 1–5 and 5–10 fixed income maturities. At the 20 year horizon the portfolio includes only 10–20y Treasuries and equity. At that stage the DBO along such scenario is very low and the current portfolio composition is the result of buying and selling decisions taken at the beginning



	0	1	2	3	5	10	20
no expire	70000.00	11498.38	4518.69	506.62	0.00	6762.64	4092.63
20-30	0.00	0.00	0.00	17886.81	0.00	0.00	0.00
10-20	0.00	0.00	0.00	0.00	0.00	1906.29	3010.14
5-10	14945.23	19178.63	17922.99	1453.12	27110.48	2602.14	0.00
1-5	15054.77	32285.54	22745.18	18185.45	295.26	0.00	0.00

Fig. 13.8 Optimal dynamic allocation time to maturity worst scenario (Y axis, thousands EUR)

of the last stage. Extraordinary sponsor's contributions are systematically avoided by selling assets, preserving the asset portfolio duration, and compounding portfolio profits to achieve a FR above 1: the introduction of a EUR -4 mln net DBO target at the 20 year horizon proves very effective even if at that point very limited resources are left. A negative net DBO at the decision horizon provides a positive *end-effect* value: from the ALM model formulation, such value does reflect the terminal ability of the PF to fund all future liabilities.

Figures 13.7 and 13.8 provide summary information on the portfolio strategy employed by the PF under a negative net DBO scenario, resulting nevertheless into a slightly positive funding surplus at T. Under any other scenario the final net DBO value improves. At the same time at the horizon, no extra resources are requested to the sponsor and the risk-adjusted return target is achieved. The latter depends directly on the assumptions on the correlation matrix introduced in the risk capital model: namely that all assets' underlying risk factors are perfectly correlated. We will provide in a separate study a range of relevant results for the cases of market or regulatory based correlation matrices.

All the analyses have been conducted relying on a simple multivariate normal asset returns model and a set of exogenous liabilities generated by the actuarial division of an occupational PF.

13.7 Conclusion

In this chapter we have presented the key elements of an ALM model formulated so as to effectively incorporate several relevant PF short-, medium- and long-term features and be consistent with recently introduced regulatory (e.g. Solvency II) and industry standards.

From a modeling viewpoint the combination along the scenario tree of decision nodes – where buying and selling decisions are allowed in the face of residual uncertainty – with intermediate nodes – from which no multiple branching originates and where pension payments and portfolio income are required to match–, allows a more flexible and effective portfolio and liquidity hedging policy without leading to a curse-of-dimensionality problem.

The adopted combination of liquidity and interest rate risk hedging in the short run, a sufficient risk-adjusted return in the medium term and long-term funding surplus goals, with the explicit inclusion of sponsor's contributions in the model proves very effective to recover a funding equilibrium for a DB pension fund. This framework is indeed at the grounds of an ongoing development of ALM methods and related tools for PF management.

Acknowledgements This research includes the formulation of a pension fund ALM problem reflecting a real-world case-study developed in cooperation between Allianz Investment Management and University of Bergamo. The presented numerical evidence has been modified and rescaled for confidentiality reasons but it does reflect actual operational conditions and the presented results remain valid.

Appendix

In Tables 13.2 and 13.3 we show the estimated statistical parameters adopted to generate correlated quarterly returns through a Monte Carlo simulation (Glasserman, 2003; Consigli et al., 2012a) for each benchmark i and node n. Once the return scenarios are aggregated in tree form, they are passed to an algebraic language deterministic model generator, to produce the stochastic program deterministic equivalent instance (Consigli and Dempster, 1998). The returns' statistics have been estimated on an historical window composed of 63 observations starting from the first quarter of 1999 till the third quarter of 2014.

Table 13.2 Average annual price (ρ) and income (ξ) returns of the entire asset universe. These parameters are estimated as the historical mean value of time series from January 1, 1999 to December 31, 2014. Data are in percentage

	ρ	ξ
EURIBOR 3m	2.41	
Treasury (T) 1–3y	−0.78	4.25
Treasury (T) 3–5y	0.13	4.38
Treasury (T) 5–7y	0.75	4.56
Treasury (T) 7–10y	1.35	4.28
Treasury (T) 10+y	1.73	4.87
Secururitized (S)	0.70	4.16
Corporate Investment Grade (IG)	0.23	4.78
Corporate High Yield (HY)	−0.93	8.05
Real Estate (R-E)	6.72	3.24
Equity (E)	6.00	3.67
TIPS 3–5y	1.53	1.74
TIPS 10+y	3.29	2.52

Table 13.3 Estimated variance-covariance matrix of the price annual returns with historical data from January 1, 1999 to December 31, 2014 of the asset universe: EURIBOR 3m (E-3m), Treasury (T), Securitized (Sec), Corporate Investment Grade (IG) and High Yield (HY), Real Estate (R-E), Equity (Eq) and TIPS

	E-3m	T 1–3y	T 3–5y	T 5–7y	T 7–10y	T 10+y	Sec	IG	HY	R-E	Eq	TIPS 3–5y	TIPS 10+y
E-3m	0.0002	0.5467	−0.0001	−0.0002	−0.0002	−0.0005	−0.0003	−0.0005	−0.0016	−0.0015	−0.0015	0.0001	−0.0004
T 1–3y	0.5467	0.0010	0.0018	0.0023	0.0026	0.0029	0.0016	0.0010	−0.0014	−0.0021	−0.0033	0.0009	0.0012
T 3–5y	−0.0001	0.0018	0.0036	0.0049	0.0059	0.0073	0.0036	0.0024	−0.0019	−0.0018	−0.0054	0.0018	0.0039
T 5–7y	−0.0002	0.0023	0.0049	0.0069	0.0086	0.0111	0.0052	0.0035	−0.0021	−0.0011	−0.0061	0.0025	0.0067
T 7–10y	−0.0002	0.0026	0.0059	0.0086	0.0111	0.0148	0.0065	0.0044	−0.0027	0.0002	−0.0066	0.0030	0.0093
T 10+y	−0.0005	0.0029	0.0073	0.0111	0.0148	0.0221	0.0087	0.0062	−0.0018	0.0049	−0.0028	0.0031	0.0145
Sec	−0.0003	0.0016	0.0036	0.0052	0.0065	0.0087	0.0046	0.0036	0.0031	0.0034	0.0013	0.0022	0.0068
IG	−0.0005	0.0010	0.0024	0.0035	0.0044	0.0062	0.0036	0.0043	0.0110	0.0073	0.0065	0.0022	0.0072
HY	−0.0016	−0.0014	−0.0019	−0.0021	−0.0027	−0.0018	0.0031	0.0110	0.0717	0.0456	0.0595	0.0040	0.0172
R-E	−0.0015	−0.0021	−0.0018	−0.0011	0.0002	0.0049	0.0034	0.0073	0.0456	0.0653	0.0614	0.0012	0.0171
Eq	−0.0015	−0.0033	−0.0054	−0.0061	−0.0066	−0.0028	0.0013	0.0065	0.0595	0.0614	0.0939	0.0001	0.0151
TIPS 3–5y	0.0001	0.0009	0.0018	0.0025	0.0030	0.0031	0.0022	0.0022	0.0040	0.0012	0.0001	0.0027	0.0051
TIPS 10+y	−0.0004	0.0012	0.0039	0.0067	0.0093	0.0145	0.0068	0.0072	0.0172	0.0171	0.0151	0.0051	0.0218

References

H. Aro, T. Pennanen, Liability driven investment in longevity risk management, in *Financial Decision Making Under Uncertainty*, ed. by G. Consigli, D. Kuhn, P. Brandimarte. International Series in Operations Research and Management Science (Springer, US, 2016), pp. 121–136

J.R. Birge, F. Louveaux, *Introduction to Stochastic Programming* (Springer, New York, 1997)

G. Consigli, M.A.H. Dempster, Dynamic stochastic programming for asset-liability management. Ann. Oper. Res. **81**, 131–162 (1998)

G. Consigli, V. Moriggia, Applying stochastic programming to insurance portfolios stress-testing. Quant. Finance Lett. **2**(1), 7–13 (2014)

G. Consigli, G. Iaquinta, V. Moriggia, Path-dependent scenario trees for multistage stochastic programmes in finance. Quant. Finan. **12**(8), 1265–1281 (2012a)

G. Consigli, G. Iaquinta, V. Moriggia, M. di Tria, D. Musitelli, Retirement planning in individual asset–liability management. IMA J. Manag. Math. **23**(4), 365–396 (2012b)

G. Consigli, D. Kuhn, P. Brandimarte, Optimal financial decision making in finance, in *Financial Decision Making Under Uncertainty*, ed. by G. Consigli, D. Kuhn, P. Brandimarte. International Series in Operations Research and Management Science (Springer, US, 2016), pp. 255–290

A. Consiglio, F. Cocco, S.A. Zenios, Scenario optimization asset and liability modelling for individual investors. Ann. Oper. Res. **191**, 152–167 (2007)

A. Consiglio, M. Tumminello, S.A. Zenios, *Designing Guarantee Options in Defined Contributions Pension Plans* (The Wharton Financial Institutions Center, Philadelphia, 2015)

M.A.H. Dempster, M. Germano, E.A. Medova, M.I. Rietbergen, F. Sandrini, M. Scrowston, Designing minimum guaranteed return funds. Quant. Finance **7**(2), 245–256 (2007)

M.A.H. Dempster, E.A. Medova, Y.S. Yong, Comparison of sampling methods for dynamic stochastic programming, in *Stochastic Optimization Methods in Finance and Energy* (Springer, New York, 2011), pp. 389–425

C.L. Dert, A dynamic model for asset liability management for defined benefit pension funds. Worldw. Asset Liabil. Model. **10**, 501–536 (1998)

J. Dupačová, G. Consigli, S.W. Wallace, Scenarios for multistage stochastic programmes. Ann. Oper. Res. **100**, 25–53 (2001)

D. Franzen, G. Scheuenstuhl, J. Yermo, S. Blome, K. Fachinger, *Pension Fund Regulation and Risk Management: Results from an ALM Optimisation Exercise*, vol. 8 (OECD Publishing, Paris, 2007)

A. Geyer, W.T. Ziemba, The Innovest Austrian pension fund financial planning model InnoALM. Oper. Res. **56**(4), 797–810 (2008)

P. Glasserman, *Monte Carlo Methods in Financial Engineering* (Springer, New York, 2003)

A.K. Konicz, D. Pisinger, K.M. Rasmussen, M. Steffensen, A combined stochastic programming and optimal control approach to personal finance and pensions. OR Spectr. **37**(3), 583–616 (2015)

E.A. Medova, J.K. Murphy, A.P. Owen, K. Rehman, Individual asset liability management. Quant. Finance **8**(6), 547–560 (2008)

J.M. Mulvey, K.D. Simsek, Z. Zhang, Improving investment performance for pension plans. J. Asset Manag. **7**(2), 93–108 (2006)

J.M. Mulvey, C. Ural, Z. Zhang, Improving performance for long-term investors: wide diversification, leverage, and overlay strategies. Quant. Finance **7**(2), 175–187 (2007)

J.M. Mulvey, K.D. Simsek, Z. Zhang, F.J. Fabozzi, W.R. Pauling, Assisting defined-benefit pension plans. Oper. Res. **56**(5), 1066–1078 (2008)

C.R. Nelson, A.F. Siegel, Parsimonious modeling of yield curves. J. Bus. **60**, 473–489 (1987)

OECD. OECD factbook 2011, public finance-pension expenditure (2011)

D. Pachamanova, N. Gulpinar, E. Canakoglu, Robust data-driven approaches to pension fund asset-liability management under uncertainty, in *Financial Decision Making Under Uncertainty*, ed. by G. Consigli, D. Kuhn, P. Brandimarte. International Series in Operations Research and Management Science (Springer, US, 2016), pp. 89–119

G.Ch. Pflug, A. Świetanowski, Dynamic asset allocation under uncertainty for pension fund management. Control Cybern. **28**, 755–777 (1999)

S. Senders, Survey of the asset and liability management practices of European pension funds. Technical report, EDHEC (2010)

C. Shaw, Fifty years of united kingdom national population projections: how accurate have they been? Popul. Trends **128**, 8 (2007)

W.T. Ziemba, J.M. Mulvey (ed.), *Worldwide Asset and Liability Modeling*, vol. 10 (Cambridge University Press, Cambridge, 1998)

Chapter 14
Currency Hedging for a Multi-national Firm

Markku Kallio, Matti Koivu, and Rudan Wang

Abstract This paper develops a multi-stage stochastic programming (SP) approach aiding a European company in currency hedging. While cash management concerns several major currencies, our pilot model deals with US$ and € only. Equilibrium correction models, Taylor rule based models and a random walk model are compared for exchange rate prediction. Risks related to exchange rate and sales forecast errors are hedged. Numerical results indicate that the current hedging policy roughly amounts to the same as no hedging at all. We demonstrate how repeated hedging activity reduces risk and thereby suggests avoid excessive risk averse behavior. Out-of-sample tests over the period 2004–2013 indicate that optimized hedging can increase net profits before taxes by about 20% over current policy. Average performance improvement of the random walk model is outperformed in terms of profit improvement by all other models we considered. Out-of-sample results show that single-stage SP yields approximately the same average improvement as multi-stage SP but the latter is more robust in terms of reduced variance.

Keywords Exchange rate • Currency hedging • Financial modeling • Stochastic programming

---

Contributions at the EqC model building stage were provided by Teemu Pennanen. The CEO and CFO of the case company are gratefully acknowledged for their support and participation. We also thank two anonymous reviewers for valuable comments and suggestions for future research.

M. Kallio (✉)
Department of Information and Service Economy, Aalto University School of Business, Helsinki, Finland
e-mail: markku.kallio@aalto.fi

M. Koivu
Nordic Investment Bank, Helsinki, Finland
e-mail: Matti.Koivu@nib.int

R. Wang
Department of Economics, Coventry University, London
e-mail: wangrudan@hotmail.com

G. Consigli et al. (eds.), *Handbook of Recent Advances in Commodity and Financial Modeling*, International Series in Operations Research & Management Science 257,
DOI 10.1007/978-3-319-61320-8_14

14.1 Introduction

Exchange rate hedging is an important problem for multi-national firms. Over the years, economics has suggested a number of theories as a basis for forecasting the exchange rates. For example, the widely known purchasing power parity theory (PPP) which links exchange rates to price levels, the flexible price monetary model (Frenkel, 1976), the sticky-price monetary model (Dornbusch, 1976) and the portfolio balance approach (Branson, 1977). These models have dominated the literature on the exchange rate during the 1970s and 1980s. More recent literature has pointed to equilibrium correction models (Engle and Granger, 1987) and to a new direction in macroeconomics by incorporating endogenous Taylor's (Taylor, 1993) rules into exchange rate models; unlike the earlier models, this more recent approach reflects how monetary policy is actually conducted or evaluated and offers a different explanation of exchange rate dynamics.

In this article we employ multi-stage stochastic programming (SP), for choosing favorable hedging strategies employing standard instruments. The approach is developed for a multi-national company[1] to aid the CFO to explore alternatives to the current hedging practice. The headquarters and major production plants of this electronics firm are located in an EU country within Eurozone. The parent company is listed on a domestic stock exchange only. There are subsidiaries elsewhere within EU as well as in North America, Brazil, Asia (e.g., China, India, South Korea, Japan) and Australia. Many of them are acquired through acquisitions. Most of the subsidiaries are only engaged in sales and services but some also in production. The firm serves customers in more than 100 countries. Financial management is centralized to parent company. The centralized cash management concerns several major currencies, € being the base currency and US$ the main foreign revenue currency. Therefore, at the present stage we deal with US$ and € only.

We develop a multi-stage SP model employing equilibrium correction models, Taylor rule based models and a random walk benchmark model for exchange rate prediction. Risks related to exchange rate and sales forecast errors are hedged. To find a favorable hedging strategy, first, alternative objectives and hedging policy constraints are considered. Numerical experiments concerning year 2004 indicate that the current practice of the firm hedging 50% of dollar revenues using forwards barely results in improvement over no hedging at all.[2] Instead, alternative optimized hedging can bring improvements. We suggest avoid excessive risk aversion by demonstrating how repeated hedging activity over months and years reduces risk in average hedging performance. Second, we provide out-of-sample tests for 2004–2013. The results show that optimized hedging can provide increase in net

[1]The CFO participated actively in the development and analysis of numerical tests of the EqC model. The optimization software was implemented using an Excel interface, a familiar tool for the CFO. For confidentiality reasons we suppress the name of the firm and modify data concerning costs and revenues.

[2]This policy was chosen by the board and communicated to the shareholders.

profits by about 20%. Average performance improvement over current policy for the random walk model is outperformed by all other models considered. Out-of-sample results show that a single-stage SP model, which is easier to understand and implement, yields approximately the same average improvement as multi-stage SP but the latter is more robust in terms of reduced variance.

The rest of the paper is organized as follows. Section 14.2 presents a brief review of exchange rate studies as well as introduces the models for exchange rate prediction which are used in our numerical analysis. In Sect. 14.3 an optimization model with restrictions for hedging policies and criteria for evaluation is developed. In Sect. 14.4 a number of numerical tests illustrate the performance of hedging strategies in terms of probability distribution of terminal asset value, the preferred criterion of the CFO for comparing alternative strategies. Most importantly, Sect. 14.5 presents out-of-sample performance comparisons of equilibrium correction models, Taylor rule based models and random walk model. Performance of each model and hedging policy is measured by terminal asset value in a 12 month out-of-sample tests. Section 14.6 concludes.

14.2 Exchange Rate Models

Following a review of exchange rate studies, this section presents models which we use to generate exchange rate scenarios for stochastic programming. The first model is an equilibrium correction model employing historical exchange rate data for prediction. The second model employs both historical data as well as current economic outlook based on macro-economic fundamentals serving as leading indicators for exchange rate predictions. The third model is a random walk model to be used as a benchmark.

14.2.1 Review of Exchange Rate Studies

The collapse of the Bretton-Woods fixed exchange-rate system in the early 1970s marked the start of the modern research on exchange-rate determination. Testing exchange rate models became popular after the major industrialized economies adopted floating exchange rates.

Studies of 1970s and early 1980s were mainly focused on asset market models (for example, Branson 1977; Frenkel 1976; Lewis 1988). Most found evidence supporting the exchange rate models of the 1970s. The landmark paper by Meese and Rogoff (1983a) challenged these findings by demonstrating that a simple random walk model performed as well as any of the tested structural exchange rate models in out-of-sample forecasting of the main US dollar exchange rates.

Since Meese and Rogoff (1983a, b) numerous studies have been devoted to developing exchange rate prediction models that could beat the random walk.

Studies by e.g. Mark (1995) and MacDonald and Taylor (1994) find that the exchange rate predictability of monetary models can be improved by extending the forecast horizon. The comprehensive study by Cheung et al. (2005) survey the literature of the 1990s and conclude that accuracy of exchange rate forecasts is very dependent on the assumptions of the data generating process and that there is no single model that would work well in all situations. Sweeney (2006) finds strong support for mean reversion in real and nominal exchange rates and demonstrates that the mean reversion models outperform random walk in out of sample forecast accuracy.

The use of Taylor rule based monetary policy functions in forecasting exchange rates have been extensively studied in recent years; see e.g. Engel and West (2005) and Molodtsova et al. (2011). Results of various studies indicate that variants of the Taylor rule perform reasonably well in exchange rate forecasting, for example, Mark (2009), Molodtsova and Papell (2009), and Ince (2014) using time series data, as well as Engel et al. (2007, 2015), Mark and Sul (2011), and Galimberti and Moura (2013) using panel data. All these papers have found that the Taylor rule based exchange rate models improve the forecasting ability of exchange rates.

Based on these findings, we utilize three different types of models for forecasting exchange rates in our hedging applications. First, the random walk model is used as the simplest possible benchmark model against which the performance of the other models will be evaluated. Second, we use a mean reverting autoregressive model (equilibrium correction model), which is a relatively simple extension of the random walk model, because Sweeney (2006) finds strong support for mean reversion models in forecasting exchange rates. Third, we employ two versions of Taylor rule based models which have been shown to work well in forecasting studies.

14.2.2 An Equilibrium Correction Model

Let Z_τ denote the (average) exchange rate (€/\$) for month τ, and define

$$x_\tau = \log(1/Z_\tau).$$

To ensure stationary time series, we deal with increments $\Delta x_\tau = x_\tau - x_{\tau-1}$. The equilibrium correction model EqC for x_τ is

$$\Delta x_{\tau+1} = A\Delta x_\tau + \alpha(x_\tau - \mu) + \epsilon_\tau, \tag{14.1}$$

where random error terms ϵ_τ are assumed independent with identical normal distributions of zero mean. Parameter $\mu = \log \eta$ is the logarithm of the equilibrium exchange rate η, and parameters A, α as well as the variance σ^2 of error ϵ_τ are subject to estimation.

Fig. 14.1 Monthly \$/€ exchange rate in 1987–2003

Using monthly data from 1987 to 2003 parameter estimates $A = 0.359$, $\alpha = -0.0272$, and $\sigma = 0.0234$ were obtained for EqC. Based on Fig. 14.1 and using management judgment, the chosen equilibrium exchange rate is $\eta = 1.10$ \$/€ is so that $\mu = \log(1.10)$ in (14.1). We analyze sensitivity of this assumption in Sects. 14.4 and 14.5

14.2.3 Taylor Rule Based Models

The link between the interest rate and the macro fundamentals originates with the central banks approach to monetary policy. In this study, we use the extended form of Taylor rule by including real exchange rate, assuming inertial hypothesis regarding the conduct of monetary policy and adopting stock price index to represent wealth effect. We consider the baseline specification for monetary policy-makers' interest rate of month τ as:

$$i_\tau = (1-\rho')(\nu + \lambda\pi_\tau + \gamma y_\tau + \kappa w_\tau + \phi q_\tau) + \rho' i_{\tau-1} + \upsilon_\tau \tag{14.2}$$

where i_τ is the actual observable short-term nominal interest rate, π_τ is the inflation rate, y_τ is the output gap (percentage deviation of actual real GDP from an estimate of its potential level), q_τ is the logarithm of the real exchange rate, w_τ is the logarithm stock price index, ρ' denotes the degree of interest rate smoothing, ν is a constant and υ_τ is the error term also known as the interest rate smoothing shock.

Model (14.2) is used for the US. Taking Germany as the benchmark country for euro region, the monetary policy reaction function for the euro is as equation (14.2) with $\phi = 0$. Parameter values and levels of economic fundamentals are country specific. We let subscript d refer to the US and e to Germany, respectively.

We assume uncovered interest rate parity (UIP) holds with rational expectations. Then given interest rates $i_{d\tau}$ and $i_{e\tau}$ at stage τ, saving 1 € at τ yields at $\tau + 1$ an amount equal to the expected yield of saving $1/Z_\tau$ \$ in a dollar account; i.e., $\exp(\delta i_{e\tau}) = \exp(\delta i_{d\tau}) E[Z_{\tau+1}]/Z_\tau$ where $\delta = 1/12$ is the time increment. Taking logarithms on both sides, recalling that $Z_\tau = \exp(-x_\tau)$ and assuming that the monthly logarithmic increment $\Delta x_{\tau+1} = x_{\tau+1} - x_\tau$ of the nominal exchange rate \$/€ is normally distributed with variance σ_T^2, independent of τ, we obtain $\delta(i_{e\tau} - i_{d\tau}) = \log E[\exp(-\Delta x_{\tau+1})] = -E[\Delta x_{\tau+1}] + \sigma_T^2/2$ where the latter equality follows from expected value formula for a log-normal random variable. Hence, $E[\Delta x_{\tau+1}] = \delta(i_{d\tau} - i_{e\tau}) + \sigma_T^2/2$, and subtracting the Taylor rule (14.2) for the US multiplied by δ from that of Germany, the exchange rate forecasting model TrE with Taylor rule fundamentals follows:

$$\begin{aligned}\Delta x_{\tau+1} = \eta_\tau &+ \psi \\ &+ \psi_{e\pi}\pi_{e\tau} + \psi_{ey} y_{e\tau} + \psi_{ew} w_{e\tau} + \rho_e i_{e,\tau-1} \\ &- \psi_{d\pi}\pi_{d\tau} - \psi_{dy} y_{d\tau} - \psi_{dw} w_{d\tau} - \rho_d i_{d,\tau-1} - \psi_q q_\tau \end{aligned} \tag{14.3}$$

where error terms η_τ are assumed i.i.d. $N(0, \sigma_T^2)$ and ψ is a constant. Suppressing subscripts e and d, the other parameters can be expressed in terms of parameters in (14.2) as follows: $\psi_\pi = \lambda(1-\rho)\delta$, $\psi_y = \gamma(1-\rho)\delta$, $\psi_w = \kappa(1-\rho)\delta$ and $\rho = \rho'\delta$ for both countries, and $\psi_q = \phi(1-\rho)\delta$ for the US. We also consider a nested model TrN which results when a component βs_τ is added to the right hand side of (14.3). Here s_τ is a trend obtained by exponential smoothing from the history of increments Δx_τ with smoothing weight 0.5. Table 14.1 shows estimated parameter values and standard deviation of the error term based on monthly data from 1993–2003.

14.2.4 Random Walk Model

For the random walk model RW the monthly log-increments of exchange rate are independent random variables with identical normal distributions of zero mean. Hence for all τ, we have

$$\Delta x_{\tau+1} = \epsilon'_\tau, \tag{14.4}$$

where random error terms ϵ'_τ have mean zero and variance σ'^2. Using monthly data from 1993 to 2003 the standard deviation is estimated to $\sigma' = 0.0266$.

Table 14.1 Parameter estimates and standard deviation of the error term for the Taylor rule based models TrE and TrN

Variable	Parameter	Estimated value	
		TrE	TrN
Inflation π_τ	$\psi_{e\pi}$	−0.000503	0.000429
	$\psi_{d\pi}$	−0.0005545	0.00227
GDP gap y_τ	ψ_{ey}	0.0005442	0.00193
	ψ_{dy}	0.000716	−0.00853
Wealth w_τ	ψ_{ew}	0.002856	0.00983
	ψ_{dw}	0.00442	0.00735
Interest rate $i_{\tau-1}$	ρ_e	−0.000419	−0.00090
	ρ_d	−0.0001190	0.00579
Real exch. rate q_τ	ψ_q	0.000910	−0.0479
Trend s_τ	β	n.a.	−0.283
Constant	ψ	0.00821	0.00826
Error η_τ	σ_T	0.0267	0.0239

14.3 A Dynamic Hedging Model

We now develop a model for multi-stage stochastic optimization of hedging strategies. First, we define the scenario tree for exchange rates. Second, we introduce sales revenues and direct production costs, and update the scenario tree to be employed in stochastic optimization. Third, we discuss currency forwards and options to be used as hedging instruments. Finally, decision variables are defined for the model, different types of hedging constraints are introduced and alternative objective functions are considered.

14.3.1 Exchange Rate Scenarios

The company employs a rolling planning and control process. Each month a forecast is produced on sales over the coming 12 month period. Consequently, a 12 month horizon is adopted to our model as well. The horizon is subdivided into a number of stages $t = 0, 1, \ldots, T$. The length T_t of period t (ending at stage t, $t = 1, \ldots, T$) in this subdivision is variable: we subdivide the 12 month time horizon into $T = 5$ time steps of $1 + 1 + 1 + 3 + 6 = 12$ months.

Based on each of the exchange rate models, randomized Sobol sequences are used to generate scenario trees. Branches per node vary by stage: for each node at stage t = 0, 1, 2, 3 and 4, respectively, there are 10, 6, 4, 4 and 2 branches. For each branch, exchange rates are generated month by month. For each month, the models (EqC, TrE, TrN and RW) employ the exchange rates generated for the preceding months. For TrE and TrN to determine the expected monthly increment we use levels of economic fundamentals observed at the initial stage $t = 0$. The

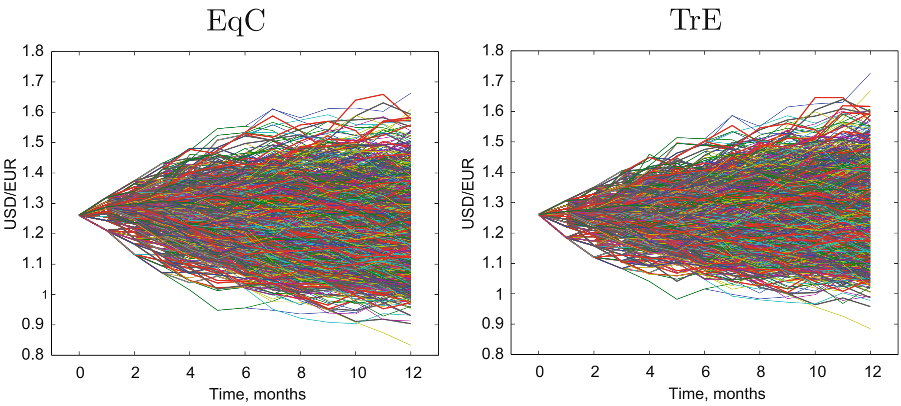


Fig. 14.2 Exchange rate scenarios ($/€) over 12 months in 2004 based on models EqC with $\eta =$ 1.1$/€ and TrE

$10 \times 6 \times 4 \times 4 \times 2 = 1920$ scenarios in the entire scenario trees for EqC and TrE in 2004 are depicted in Fig. 14.2.

14.3.2 Revenues and Expenditures

Let R^f_τ and R^a_τ denote the monthly US$ sales forecast and the actual monthly sales, respectively, for month τ. The logarithmic monthly forecast error is then given by

$$R^e_\tau = \ln R^a_\tau - \ln R^f_\tau. \tag{14.5}$$

The monthly values of R^e_τ during 2001–2003 are displayed in Fig. 14.3.

Annual US$ sales forecasts are made once a year. As can be seen from Fig. 14.3, the forecast error has zero mean on average and longer horizon forecast errors are not significantly larger than the short horizon forecast errors. Based on these observations we model the development of the sales error with a time series model

$$\Delta R^e_\tau = \alpha_e R^e_{\tau-1} + \epsilon_\tau, \quad \epsilon_\tau \sim N(0, \sigma_e). \tag{14.6}$$

We estimate the model for the sales error using the data displayed in Fig. 14.3. The estimated parameter values are $\alpha_e = -0.757$ and $\sigma_e = 0.225$.

We assume that sales over nearest three months is known with certainty so that $R_\tau = R^a_\tau$, for $\tau \leq 3$. For $\tau > 3$, using the monthly US$ sales forecast for the year 2004, the stochastic US$ sales R_τ is given by

$$R_\tau = R^f_\tau e^{R^e_\tau}. \tag{14.7}$$

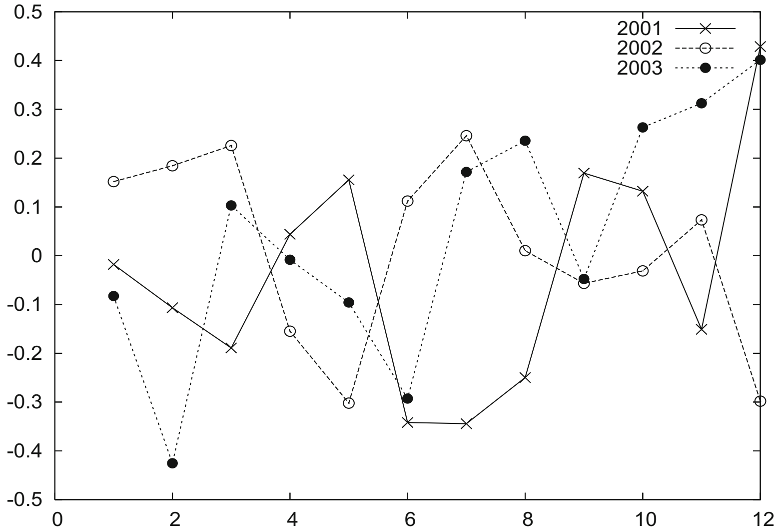


Fig. 14.3 The logarithmic monthly forecast error for years 2001–2003

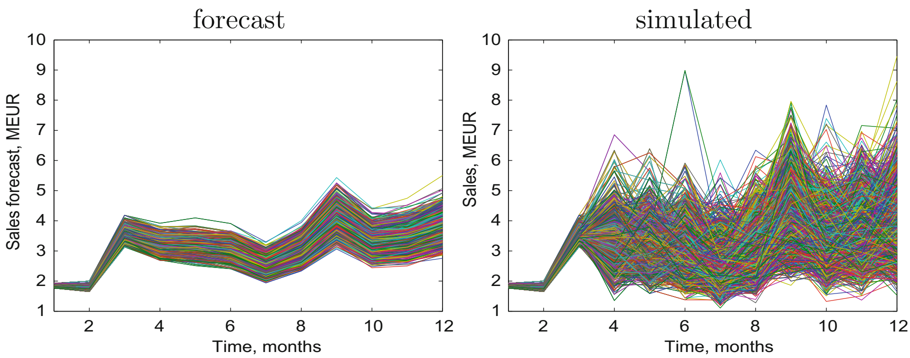


Fig. 14.4 Scenarios of a monthly US$ sales forecast in M€ (*left*) and of simulated monthly US$ sales in M€ (*right*) for 2004 using exchange rate model EqC with $\eta = 1.1\$/€$

Based on Fig. 14.3, the forecast error of December 2003 is 0.4. Hence, when (14.6) is applied for January 2004 with $\tau = 1$ we have $R_0^e = 0.4$. Using exchange rate model EqC with $\eta = 1.1\$/€$, Fig. 14.4 (left) illustrates the simulated scenarios of sales forecast in M€ based on a deterministic monthly US$ sales forecast for 2004. In Fig. 14.4 (right) simulated monthly US$ sales based on (14.6) and (14.7) are converted to M€ using simulated exchange rates. Approximately 30% of the uncertainty in € sales is caused by the exchange rate fluctuations while the majority of the uncertainty results from the sales forecast errors.

With a profit margin $1 - c$, the euro expenditure of production (M€) in month τ is

$$E_\tau = cR_\tau Z_{\tau-\omega} \tag{14.8}$$

where ω is the delay between production and delivery. Hence, the profit of month τ in € is $Z_\tau R_\tau - E_\tau = R_\tau(Z_\tau - cZ_{\tau-\omega})$. For numerical tests we use $c = 0.55$ and $\omega = 3$ months. To construct a scenario tree for stochastic optimization, a natural approach would be to construct an independent scenario tree for dollar sales revenues and adopt the product tree of exchange rates and revenues for optimization. However, even with a small number of scenarios for revenues, the product tree quickly becomes too large for our computing facilities. To cope with the curse of dimensionality, we adopt an approach where the scenario tree is constructed by sampling realizations for the USD/EURO exchange rate and sales forecast error from a two dimensional randomized Sobol sequence. For each arc joining stages t and $t+1$ in the tree for exchange rates, we generate realizations of sales R_τ and costs E_τ for the months τ of period $t+1$. Thereby we obtain 1920 scenarios for revenues and costs as well. We believe this is sufficient to take into account sales forecast errors.

The risks to be hedged concern dollar revenues R_t and costs E_t which are obtained summing up monthly revenues and costs of months τ in period t. Given stochastic revenue and cost streams $\{R_t\}$ and $\{E_t\}$, a hedging strategy is chosen over 12 months at stage $t = 0$.

14.3.3 Forwards and Options

For numerical tests we use simplifying assumptions for interest rates, forward rates and option prices which for our purposes adequately approximate market rates and prices. For the nominal interest rates, let r_{ets} and r_{dts} be the euro and dollar interest return, respectively, from time t to time s. The interest rate curves are assumed flat and the rates are taken as observed three month rates prevailing at the beginning of the year.

For $t = 0, 1, \ldots, T-1$, consider currency contracts made at stage t and maturing at stage $s > t$. The forward rate F_{ts} in €/\$ is (see Hull 2002)

$$F_{ts} = Z_t \exp(r_{ets} - r_{dts}). \tag{14.9}$$

Let $X_{ts} = (X_{xts})$ be the vector of exercise prices in €/\$ for options. Here indices x refer to alternative exercise prices. We use exercise rates which are 90, 100 and 110% of the current exchange rate Z_t; hence, $X_{ts} = (X_{xts}) = (0.9, 1.0, 1.1) \times Z_t$. We approximate the option prices using theoretical prices of the Garman-Kohlhagen model (see e.g., Hull 2002). Then the price of a European put option with an exercise price X_{xts}, is

$$P_{xts} = [X_{xts}N(-d_2) - F_{ts}N(-d_1)] \exp(-r_{ets}). \tag{14.10}$$

and the price of a European call option is

$$C_{xts} = [F_{ts}N(d_1) - X_{xts}N(d_2)]\exp(-r_{ets}), \tag{14.11}$$

where $N(\cdot)$ is the cumulative probability distribution function of the standard normal distribution, and $d_1 = d_2 + \sigma_{ts} = [\log(F_{ts}/X_{xts}) + \sigma_{ts}^2/2]/\sigma_{ts}$ with σ_{ts} being the exchange rate volatility over the period from t to s.

The proportional transaction cost is $T_f = 0.001$ for forwards and $T_o = 0.01$ for options as suggested by the CFO. Currency exchange costs are assumed zero. Estimated annual exchange rate volatility is $\sigma_{0T} = 0.092$. For numerical analysis in Sects. 14.4 and 14.5 we assume that no derivatives acquired before the initial stage $t = 0$ remain to mature during the subsequent 12-month period.

14.3.4 Model Formulation

To state the optimization model, we introduce (i) accounting and hedging variables, (ii) hedging policy constraints and (iii) alternative hedging objectives; for applications of SP, see e.g., Ziemba and Mulvey (1998).

(i) *Accounting and hedging variables*. Choice of currency portfolios, including long position in each currency, as well as choice of portfolios of derivatives determine a hedging strategy. No hedging takes place at the terminal stage T. Hence, we define the following decision variables for the currency portfolios, for $t = 0, \ldots, T-1$: e_t is cash (M€) and d_t is cash (M\$) at stage t after hedging. Initial currency positions at $t = 0$ are $e^0 = 10$ M€ and $d^0 = 0$ M\$. The amounts of forward contracts made at stage t, maturing at s, and options with exercise prices X_{xst}, for $t = 0, \ldots, T-1$ and $s > t$, are as follows: s_{ts} is forward contracts to sell dollars (M\$), b_{ts} is forward contracts to buy dollars (M\$), p_{xts} is European put options (M\$) and c_{xts} is European call options (M\$). Additionally, for all $t = 0, 1, \ldots, T$ we define accounting variables w_t (M€) such that w_t is the total asset value prior to hedging actions at stage t. Hence, initially $w_0 = c^0 + Z_0 d^0$ is a constant. For each stage $t > 0$, depending on the state of the currency market (in particular nodes of the scenario tree), different values may be chosen for the variables, and therefore, the variables c_t, d_t, w_t, s_{ts}, b_{ts}, p_{xts} and c_{xts} are stochastic.

(ii) *Hedging constraints*. The model constraints include policy based simple bounds (upper and lower limits on decision variables) and other restrictions concerning the use of derivatives to be employed as well as accounting equations for assets.

Simple bounds on variables account for possible policies on positions in each currency and on quantities of derivatives employed. For instance, short positions in options may be prohibited or use of some instruments may be excluded; see numerical illustrations in Sect. 14.4 where all lower limits are set to zero, thus prohibiting short positions. The use of an instrument may also be fixed to a given level, to zero, for instance.

There are additional policy restrictions on contracts for selling dollars (i.e., on forwards and puts). In particular, the total quantity of contracts maturing at time t are limited by revenue R_t. Let indices x refer to exercise prices of options. Then

$$\sum_{s<t}(s_{st} + \sum_{x} p_{xst}) \leq R_t \ \text{ for } t = 1, \ldots, T. \tag{14.12}$$

Similarly, we restrict contracts for buying dollars (i.e., forwards and calls). In numerical tests such upper limit is set to revenues R_t as well.

The accounting equations concern the cash balance at stage t, for $t = 0, 1, \ldots, T-1$, and the dynamics of the total value w_t, for $t = 0, 1, \ldots, T$. Recalling the initial value $w_0 = c^0 + Z_0 d^0$, the total cash value plus expenditures in options and transaction costs is equal to the total asset value w_t for all t; i.e., cash e_t (M€) satisfies the following budget balance requirement for all t

$$e_t + Z_t d_t + \sum_{s>t}[T_f(s_{ts} + b_{ts}) + \sum_{x}(1 + T_o)(C_{xts}c_{xts} + P_{xts}p_{xts})] = w_t.$$

Given euro and dollar returns r_{ets} and r_{dts} from time t to time s, the total asset value w_t at $t > 0$ prior to hedging is determined by cash at the beginning of period t, interests from the period t, revenues and expenditures of period t, as well as profits from currency derivatives acquired in stages $s < t$. Hence, the total asset value w_t (M€) prior to hedging actions at stage t is given by

$$\begin{aligned} w_t = {} & (1 + r_{e,t-1,t})e_{t-1} + Z_t[(1 + r_{d,t-1,t})d_{t-1} + R_t] - E_t \\ & + \sum_{s<t}(F_{st} - Z_t)(s_{st} - b_{st}) + \sum_{x,s<t}[\max(0, X_{xst} - Z_t)p_{xst} \\ & + \max(0, Z_t - X_{xst})c_{xst}] \end{aligned}$$

(iii) *Hedging objectives*. Our aim is to explore alternatives for risk profiles. Therefore, we consider several objective functions, each one being determined by the total asset value w_T at terminal stage T. First, we state the standard expected utility criterion employing a negative exponential utility function. If the risk aversion parameter approaches zero or infinity we obtain the expected value criterion or the worst case criterion, respectively. Finally, the weighted sum of the expected value and the worst case is proposed. In practical applications, the chosen hedging objective aims to reflect true preferences of the decision maker, such as the CFO of the firm.[3] However, as in Sect. 14.4, different objectives may be explored for learning purposes in a decision support process. – Below, $E[\cdot]$ refers to the expectation operator.

[3] Preferences of the CFO are tied to the incentive system, which for the CFO and other senior staff at the headquarters is based 1- and 3-year profit and growth of the firm.

An expected utility criterion of a risk averse decision maker is to

$$\max\ E[-\exp(-\gamma w_T)]. \tag{14.13}$$

where $\gamma > 0$ is the Arrow-Pratt absolute risk aversion coefficient. This criterion aims to balance good and poor levels of w_T. The utility function may be interpreted as a penalty which increases progressively as the terminal value decreases. Alternatively, the marginal payoff decreases with increasing terminal value.

For small γ, (14.13) approaches the expected terminal wealth criterion to

$$\max\ E[w_T]. \tag{14.14}$$

This criterion is attributed to a risk neutral decision maker. However, also for a risk averse person the expected value can be an appropriate criterion if similar decision problems repeat over time.[4]

As γ increases without limit, (14.13) approaches the worst case criterion to $\max\ \{\, l \mid l \le w_T \,\}$. In addition to the standard criteria above, we consider a weighted sum of the expected value and the worst case using a weight coefficient λ with $0 < \lambda < 1$:

$$\max\ \{\, \lambda E[w_T] + (1-\lambda) l \mid l \le w_T \,\}. \tag{14.15}$$

14.4 Comparison of Hedging Strategies Over a Single Year

The primary concern of the CFO is the probability distribution of the asset value w_T at the end of the 12 month period. This represents a risk and return profile where preferences are subjectively formed based on the distributions of w_T. For this reason, we consider a number of runs with the model exploring alternatives for such distribution for year 2004. Case 0 is without hedging. Case 1 mimics the current hedging practice. It is advanced in Case 2 relaxing previous hedging restrictions and using expected utility maximization. Speculative hedging with expected value maximization is presented in Case 3 and maximization of a weighted average of the worst case and the expected value is the objective in Case 4. We discuss the results

[4] As a simple example, consider two alternative choices A and B. Alternative A is risk less with $w_T = 50$ M€. In case B there is an 80% chance for $w_T = 60$ M€ and a 20% chance for $w_T = 40$ M€ with an expected value $E[w_T] = 56$ M€. If the choice is not repeated, a risk averse manager may prefer A to B. If such identical choices are repeated in n consecutive years with independent outcomes, then the expected terminal value for B is at 56 M€ per annum and for A the average is 50 M€, for all n. However, the probability of the annual average in case B being less than 50 M€ is only 4.0%, 2.7% and 0.6%, for n = 2, 4 and 10, respectively. Therefore, a person preferring A to B in a single trial might prefer B to A in repeated trials.

basd on $E[w_T]$ and the pdf of w_T. Expected hedging levels and asset values by time stage for Cases 0–2 are shown in Table 14.4 of the Appendix.

Case 0: No hedging. Consider the case without derivatives and \$ positions fixed to zero. The distribution of the terminal value w_T is shown in Fig. 14.5 (blue curves) using exchange rate models EqC (left) and TrE (right). Expected terminal value $E[w_T]$ is 27.5 M€ for EqC and is 26.9 M€ for TrE.

Case 1: Current hedging policy. This case mimics the current hedging practice: We prohibit both put and call options as well as forwards to buy dollars. Cash in dollars is kept at zero. Hence, we restrict cash d_t, forwards b_{st} put options p_{xst} and call options c_{xst} to zero. For $t \leq 3$, 50% of dollar revenues R_t are hedged using forwards. For $t > 3$, we require that at most 50% of dollar revenues are hedged. Furthermore, revenues at stages T and $T - 1$ can only be hedged at the preceding stages $T - 1$ and $T - 2$, respectively. Because revenues at stage 3, for instance, can be alternatively hedged at stages 0, 1 and 2, there is some freedom to choose. Therefore, we choose the hedging strategy employing the expected terminal value criterion (14.14).

Figure 14.5 (top, red curves) shows the distribution of the terminal value w_T with EqC (left) and TrE (right). We observe that the distributions of Cases 0 and 1 quite similar, and expected terminal value w_T increases from Case 0 by 0.1 M€ only. Such similarity may be explained by the fact that an unhedged dollar amount in Case 0 yields a loss if the dollar weakens and a gain in the opposite case compared with Case 1. When we take into account all scenarios and all dollar revenues over time, such gains and losses tend to cancel each other, and consequently at the end at stage T, the outcome in Case 0 without hedging is approximately the same as in Case 1 employing hedging with forwards.

Case 2: Optimal hedging with expected utility. Case 1 is advanced by allowing (but not requiring) all dollar revenues R_t be hedged with forwards s_{st} and puts p_{st}, and forwards b_{st} and calls c_{st} to buy are available as well. Hence, in this case constraint (14.12) applies to selling dollars. Similarly, buying dollars using forwards and calls is limited to R_t. Expected utility criterion (14.13) is used with risk aversion coefficient $\gamma = 0.5$/M€. Figure 14.5 (second row, red curves) show the distributions of w_T with EqC (left) and TrE (right). Expected terminal value w_T increases from Case 0 by 0.9 M€ for EqC and 0.7 M€ for TrN.

Case 3: Speculative hedging. Next, we consider an increasingly speculative hedging strategy. We take Case 2 to begin with allowing all hedging instruments to be employed and use criterion (14.14) of maximizing expected terminal value w_T. Figure 14.5 (third row, red curves) show the distributions of w_T with EqC (left) and TrE (right). Expected terminal value w_T increases from Case 0 by 1.5 M€ for EqC and 0.9 M€ for TrN.

As discussed in Sect. 14.3.4, expected value maximization can be an appropriate criterion if similar decision problems repeat over time. To illustrate, consider two and five independent 12 month hedging problems, each one being identical to the case above. If the optimal hedging strategy for a single year is adopted in each year, then the distribution of the average terminal value per annum has the same expected value as above. However, the variance decreases. Figure 14.6 shows the results of a

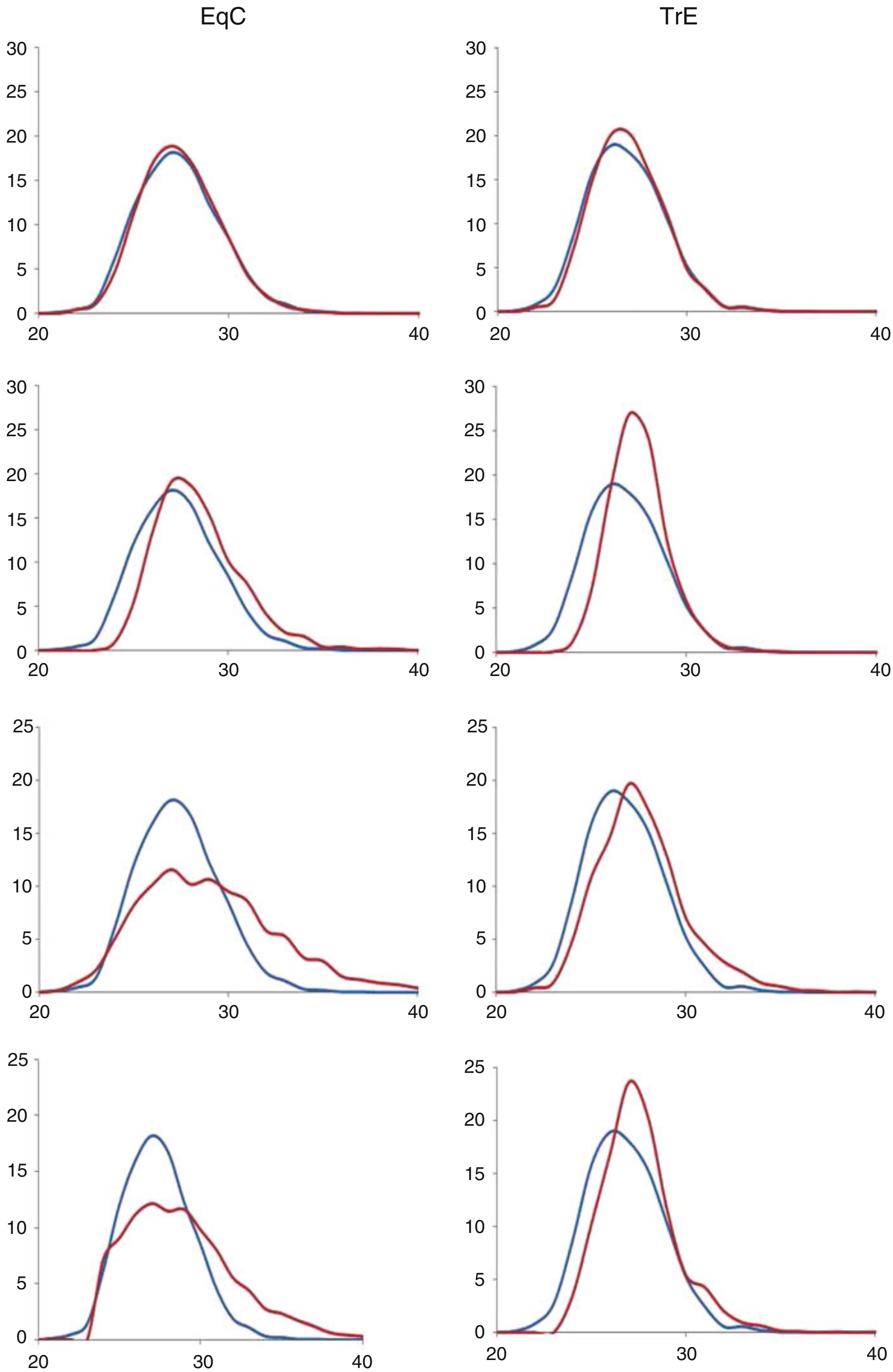


Fig. 14.5 Probability distribution of terminal assets value w_T (M€) with exchange rate models EqC (*left figures*) and TrE (*right figures*). In each figure Case 0 is depicted in *blue* and the *red curve* refers to Case 1 (*first row*), Case 2 (*second row*), Case 3 (*third row*), Case 4 (*bottom row*)

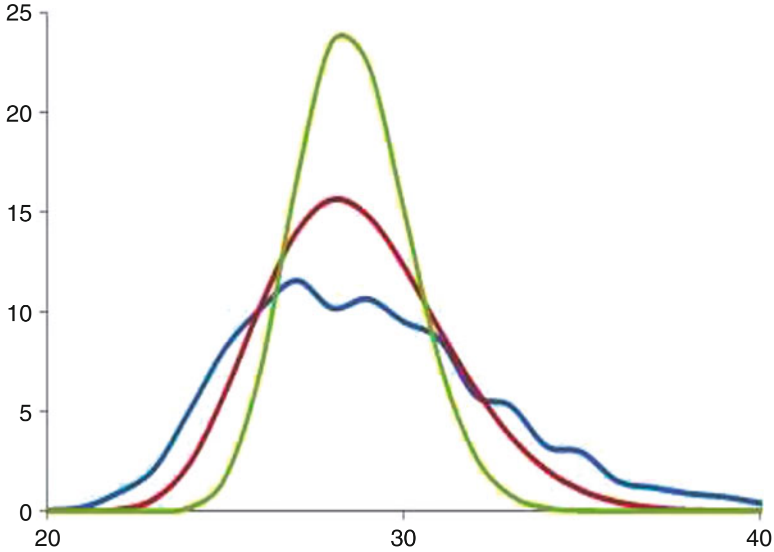


Fig. 14.6 Probability distribution of terminal asset value w_T (M€) in Case 3. Results are shown both for a single year (*blue*), as well as for two (*red*) and for five (*green*) independent repetitions

single year (blue curve) as well as for its two and five independent repetitions (red and green curves, respectively).

Case 4: Weighted objective. This case is obtained from Case 2 by switching to criterion (14.15), where the objective is to maximize the weighted average of the expected value and the worst case of the terminal value w_T. We use equal weights so that $\lambda = 0.5$ in (14.15). Figure 14.5 (bottom row, red curves) show the distributions of w_T with EqC (left) and TrE (right). Expected terminal value w_T increases from Case 0 by 1.4 M€ for EqC and 0.7 M€ for TrN.

14.5 Out-of-Sample Tests

In this section we compare EqC models with alternative equilibrium rates η, Taylor rule based models TrE and TrN, and the random walk model RW using out-of-sample tests for 2004–2013. In each test we consider three hedging alternatives: Case 0 (without hedging), Case 1 (current hedging policy) and Case 2 (optimized hedging using expected utility maximization in (14.13) with risk aversion coefficient γ to be specified below).

Out-of-sample tests are carried out separately for 12-month periods starting at 1/2004, 7/2004, 1/2005, ..., until 1/2013. The beginning of the first month of each period is denoted by t_0 for $t_0 = 0, 6, 12, \ldots, t_{max}$, where $t_{max} = 108$. Hence, $t_0 = 0$ refers to the beginning of January 2004, $t_0 = 6$ refers to the beginning of July 2004, etc. For the tests we assume rolling levels of sales forecasts as well as actual sales. Hence for any month τ, we have sales forecast $R^f_\tau = R^f_{\tau-12}$ and actual sales $R^a_\tau = R^a_{\tau-12}$. For each of the $1 + t_{max}$ 12-month periods, we assume that the initial asset value is 10 M€ and there are no derivatives to be taken into account from the past, similarly as we assumed in Sect. 14.4 at the beginning of 2004.

For each $t_0 = 0, 6, \dots, t_{max}$ we solve 12 optimization tasks over a period of 12 months starting at $t_0 + \tau$, for $\tau = 0, 1, 2, \dots, 11$ as follows. For each $t_0 + \tau$, only derivative contracts chosen by the optimal solution at the root node (at $t_0 + \tau$) are accepted for implementation. Given t_0 and τ, we observe the realized exchange rates, sales revenues and economic fundamentals up to and including stage $t_0 + \tau$, and set up the scenario tree for exchange rates, revenues and costs. For EqC we employ the most recent history of exchange rates. For TrE and TrN we employ additionally the realized levels of economic fundamentals at $t_0 + \tau$ and these levels are applied to the entire 12 month period starting at $t_0 + \tau$. Observed interest rates at $t_0 + \tau$ are adopted and forward rates, option prices as well as option exercise prices are updated. For $\tau > 0$, the initial asset value w_0 at stage $t_0 + \tau$ is determined by preceding optimization starting at $t_0 + \tau - 1$ and observed realizations of exchange rate and sales at $t_0 + \tau$. For $\tau > 0$, in the optimization over 12-month period starting at $t_0 + \tau$ we take as given the derivative contracts which were accepted already at $t_0 + \tau'$, for all $0 \le \tau' < \tau$, and which mature after stage $t_0 + \tau$. Recall that only derivative contracts chosen in the root node at stage $t_0 + \tau'$ are accepted. Figure 14.7 illustrates the models TrN and EqC with $\eta = 1.1$ \$/€ for $t_0 = 0$; i.e., for the year 2004.

For each t_0, after solving the 12 optimization problems over one year starting at $t_0 + \tau$, $\tau = 0, 1, \dots, 11$ we calculate the asset value at $t_0 + 12$, at the end of the year starting at stage t_0. This terminal value V is the sum of the realized cash value at $t_0 + 12$ and the value of those derivative contracts which are accepted at t_0, $t_0 + 1, \dots, t_0 + 11$ but not matured by $t_0 + 12$. Terminal value V is the out-of-sample performance indicator of hedging strategies. We denote by V_0, V_1 and V_2 terminal

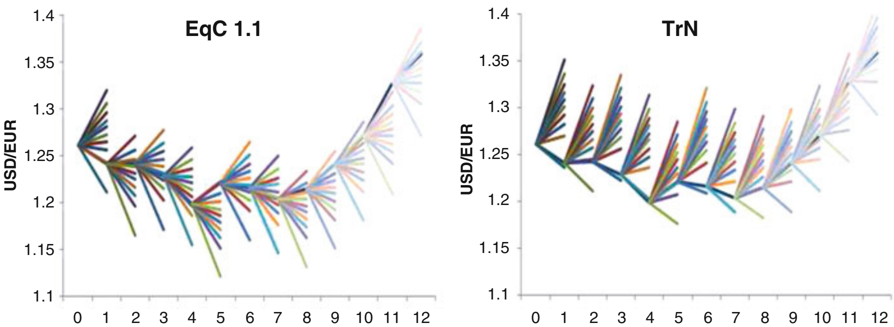


Fig. 14.7 Exchange rate scenarios in 2004 of out-of-sample simulation with $t_0 = 0$ over the first month with initial stages $t_0 + \tau$, for $\tau = 0, 1, 2, \dots, 11$, using EqC with $\eta = 1.1$ \$/€ and TrN

Table 14.2 Out-of-sample test results for 2004–2013 using multi-stage and single-stage optimization. Average terminal values (M€): V_0 is without hedging, V_1 is based on current hedging policy, V_2 is based on optimized hedging ($\gamma = 0.5$/M€); std is standard deviation and t-stat the t-statistic of $V_2 - V_1$ in 19 tests

	Model							
	EqC1.3	EqC1.2	EqC1.1	EqC1.0	EqC0.9	TrE	TrN	RW
Multi-stage optimization								
$V_2 - V_0$	1.00	1.04	0.84	0.81	0.80	0.49	0.71	0.29
$V_1 - V_0$	0.06	0.04	-0.01	0.02	0.03	0.02	0.09	0.02
$V_2 - V_1$	0.94	1.00	0.84	0.80	0.77	0.47	0.62	0.28
std	1.04	1.29	1.61	1.09	0.86	0.82	2.17	0.65
t-st	3.93	3.38	2.29	3.17	3.87	2.47	1.25	1.84
Single-stage optimization								
$V_2 - V_0$	1.02	1.03	0.96	0.82	0.84	0.59	0.65	0.49
$V_1 - V_0$	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
$V_2 - V_1$	0.93	0.94	0.87	0.73	0.76	0.50	0.57	0.40
std	1.19	1.36	1.84	1.16	1.04	1.08	2.04	0.96
t-st	3.41	3.02	2.07	2.77	3.17	2.02	1.21	1.83

values with hedging strategies of Case 0 (unhedged), Case 1 (current policy) and Case 2 (optimized policy using expected utility maximization with risk aversion coefficient shown by tables of results), respectively. Values V_1 and V_2 depend on the exchange rate model as well and the results are shown in Tables 14.2 and 14.3.

Out-of-sample results for the period 2004–2013 are shown in Table 14.2 (top). The tests indicate that optimized hedging can provide increase in net profits over current policy[5] by 6–23% – a statistically significant improvement for all except one model. Taylor rule based model TrN yielding a 14% performance improvement suffers from large variance, and thereby, its improvement lacks statistical significance. The base case model EqC1.1 yields an average improvement of 19%; however, judging from Figs. 14.1 and 14.8, EqC1.2 might have been chosen equally well, in which case the improvement is 23% in profits.[6] Improvement due to current policy over unhedged case is insignificant. Average performance improvement over current policy for the RW model is outperformed by all other models considered. Figure 14.9 shows the improvement $V_2 - V_1$ over time of optimized hedging for models EqC1.1, EqC1.2, TrE and RW.

An interesting question is to analyze what additional value does multi-stage stochastic optimization bring compared to single-stage optimization where hedging takes place at the root node only. For potential users of hedging models, the single-stage approach would be much easier to understand and implement. A multi-stage

[5]The estimate is based on 10% profit margin without hedging.

[6]Further tests with EqC1.1 indicate that the autoregressive component in (14.1) is more important than the mean reverting component; however, for the best result both components are needed.

Table 14.3 Sensitivity analysis of out-of-sample test in 2004–2007. Average results using single-stage optimization (top), moving risk aversion coefficient up and down by factor 2 (middle), and doubling the number of branches per node at stage $t = 4$ resulting in 3840 scenarios. Average terminal values (M€): V_0 is without hedging, V_1 is based on current hedging policy, V_2 is based on optimized hedging; std is standard deviation and t-stat the t-statistic of $V_2 - V_1$ in 7 tests

	Model							
	EqC1.3	EqC1.2	EqC1.1	EqC1.0	EqC0.9	TrE	TrN	RW
Multi-stage optimization, γ =0.5/M€								
$V_2 - V_0$	0.88	0.73	0.42	0.55	0.68	0.46	1.49	0.56
$V_1 - V_0$	0.16	0.15	0.11	0.14	0.13	0.07	0.24	0.07
$V_2 - V_1$	0.72	0.58	0.31	0.41	0.56	0.38	1.25	0.48
std	0.99	0.69	0.68	0.54	0.74	0.73	2.06	0.58
t-st	1.92	2.23	1.22	2.05	1.99	1.40	1.60	2.19
Single-stage optimization, γ =0.5/M€								
$V_2 - V_0$	1.03	0.81	0.41	0.59	0.80	0.66	1.39	0.80
$V_1 - V_0$	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24
$V_2 - V_1$	0.78	0.56	0.17	0.35	0.56	0.42	1.15	0.55
std	1.24	0.92	0.71	0.72	0.94	1.09	1.91	0.92
t-st	1.67	1.62	0.63	1.30	1.58	1.02	1.59	1.60
Risk aversion coefficient γ =1.0/M€								
$V_2 - V_0$	0.99	0.87	0.65	0.74	0.85	0.71	1.73	0.77
$V_1 - V_0$	0.16	0.15	0.11	0.14	0.13	0.07	0.24	0.07
$V_2 - V_1$	0.83	0.72	0.54	0.60	0.72	0.64	1.49	0.70
std	1.12	0.89	0.64	0.81	0.96	1.10	2.28	1.03
t-st	1.95	2.15	2.26	1.98	1.98	1.53	1.73	1.79
Risk aversion coefficient γ =0.25/M€								
$V_2 - V_0$	0.78	0.41	−0.06	0.11	0.33	0.04	1.50	0.07
$V_1 - V_0$	0.16	0.15	0.11	0.14	0.13	0.07	0.24	0.07
$V_2 - V_1$	0.61	0.26	−0.17	−0.03	0.20	−0.03	1.26	−0.01
std	0.75	0.71	1.39	0.85	0.49	0.30	2.07	0.14
t-st	2.17	0.98	−0.33	−0.08	1.07	−0.30	1.61	−0.13
Increased number of scenarios, γ =0.5/M€								
$V_2 - V_0$	0.85	0.70	0.33	0.52	0.68	0.43	1.52	0.49
$V_1 - V_0$	0.16	0.15	0.11	0.14	0.13	0.07	0.24	0.07
$V_2 - V_1$	0.69	0.55	0.23	0.38	0.56	0.36	1.28	0.42
std	0.95	0.70	0.78	0.52	0.74	0.68	2.06	0.50
t-st	1.92	2.10	0.77	1.96	1.99	1.39	1.64	2.21

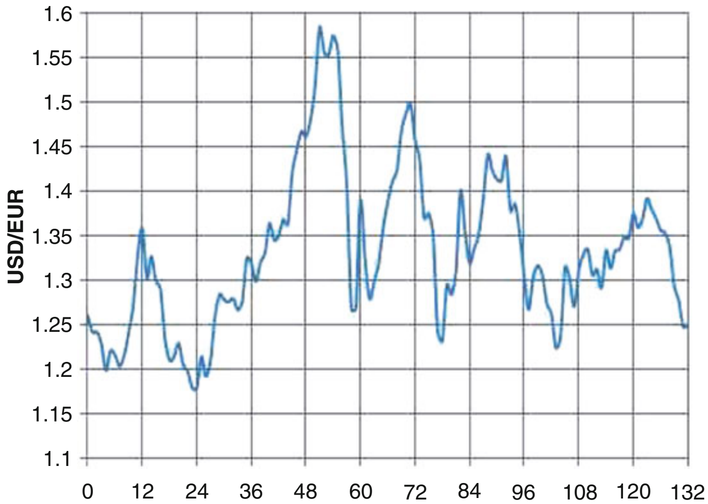


Fig. 14.8 Monthly \$/€ exchange rate 2004–2014. *Horizontal axis* shows months t_0 from the beginning of January 2004

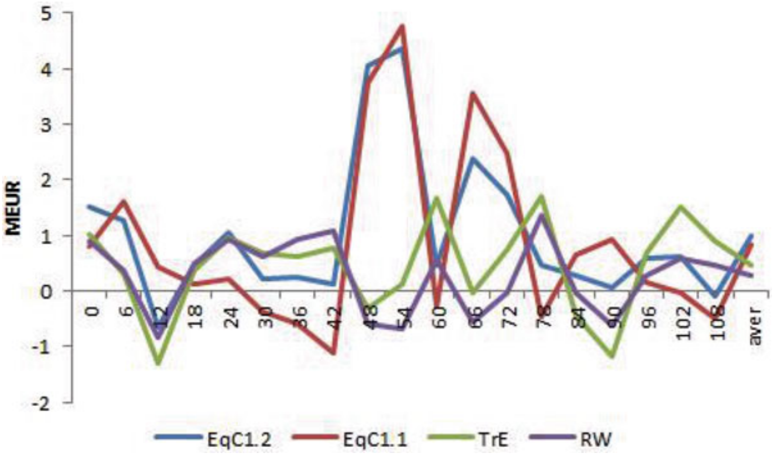


Fig. 14.9 Difference $V_2 - V_1$ (M€ per annum) in out of sample simulations for 2004–2013. Terminal value V_1 is based on current hedging policy, V_2 is based on optimized hedging ($\gamma = 0.5$/M€). *Horizontal axis* t_0 is months from the beginning of January 2004

formulation is justified if it produces improvement over single-stage model in out-of-sample tests. To test this question, we use our multi-stage model and restrict all hedging activities to take place at the root node. Thereby we obtain a single-stage formulation. Table 14.2 (bottom) shows that a single-stage SP model, yields for 2004–2013 approximately the same average improvement as multi-stage SP, but the multi-stage approach is more robust in terms of reduced variance. However, most of average performance improvements using a single-stage model are statistically significant.

For sensitivity analysis we study the period 2004–2007 as well. The results are summarized in Table 14.3. On the top it shows that optimized hedging can provide increase in net profits by 7–29% over current policy – an improvement which is statistically significant for half of the models. However, a Taylor rule based model TrN yielding the highest average performance improvement suffers from the worst variance, and thereby, it lacks statistical significance. The base case model EqC1.1 of Sect. 14.2.2 yields the lowest average improvement of 7%; however, judging from Figs. 14.1 and 14.8, EqC1.2 might have been chosen as the base case, in which case the improvement is 14% in profits.

The results for single-stage formulation are shown next in Table 14.3 (second block). Again, it shows that a single-stage SP model, yields approximately the same average improvement as multi-stage SP but the latter is more robust in terms of reduced variance. None of average performance improvements using a single-stage model is statistically significant.

Table 14.3 (middle) shows sensitivity analysis concerning the risk aversion coefficient. An increase in risk aversion coefficient γ by factor 2 appears to improve the hedging performance while a decrease by the same factor does the opposite.

One suspectible issue with our multi-stage model is the potentially significant approximation errors when presenting the exchange rate and sales uncertainty with only two scenarios (per node) in the later stage in the scenario tree. For this reason we tested the case where branches per node were doubled at each node in stages $t = 4$, thereby resulting in 3840 scenarios. Otherwise the test follows the assumptions underlying results in Table 14.3 (bottom) shows that doubling the number of scenarios does not result in significant changes in hedging performance.

14.6 Conclusions

In this article we employ multi-stage stochastic optimization for choosing favorable hedging strategies employing standard instruments. The approach is developed for a European company to aid the CFO to explore alternatives to the current hedging practice. In this company US$ is the main foreign revenue currency wherefore we deal with US$ and € only.

We develop a multi-stage SP (stochastic programming) model employing equilibrium correction models, Taylor rule based models and a random walk benchmark model for exchange rate prediction. Risks related to exchange rate and sales forecast errors are hedged. In numerical tests preferences of alternative hedging strategies is based on the terminal asset value. To find a favorable hedging strategy, first, alternative objectives and hedging policy constraints are considered. Numerical experiments indicate that the current practice of the firm hedging 50% of dollar revenues using forwards does not provide significant improvement over no hedging at all. We suggest avoid excessive risk aversion by demonstrating how repeated hedging activity over months and years reduces risk in average hedging performance. Second, we provide out-of-sample tests for 2004–2013. The results show that optimized hedging can provide increase in net profits over current hedging practice by about 20%. Average performance improvement over current policy for the random walk model is outperformed by all other models considered. Out-of-sample results show that a single-stage SP model yields approximately the same average improvement as multi-stage SP but the latter is more robust in terms of reduced variance. Because the single-stage SP model is easier to understand and implement, the user might begin with a single-stage SP.

The methodology of this study may be extended to cash management in different multinational group structures with central decision making on hedging, for example, with a holding company organizing cash pooling and risk compensation system between the different subsidiaries. For multi-national companies, there is a variety of internal ownership structures serving for different purposes. Extensions employing alternative hedging policies or aiming to fair taxation of the company are also possible. While this article deals with operational risk, future research might focus on long term financial risk, on investments abroad – financed, for instance, is USD – and on consolidation risk.

Future research tasks might also include further exploration of Taylor rule like models, which include macro indicators as predictors for exchange rates development. Of course, multi-currency hedging would be a subject of interest. In the spirit of ongoing popular discussion on the use of "big data", we might consider use of the vastly increasing supply of digital data, both numeric and non-numeric.

Finally, we raise an important question: Can SP help the CFO? In view of our out-of-sample tests, the answer is likely to be positive. However, there are several important issues to be considered prior to adoption of SP in practice. Some of them are as follows: First, is there an acceptable exchange rate model? Second, what is an acceptable level of speculation both from the point of view of the firm and the CFO. Third, how communicate the SP approach to the CFO, the CEO and the board of the firm?

Appendix: Hedging Results with Exchange Rate Model EqC and TrE

Table 14.4 Expected levels of contracts and assets in Cases 0–2. For contracts in (M\$), s_{ts} and b_{ts} are forwards to sell and buy; p_{xts} and c_{xts} are put and call options. For assets, e_t is cash (M€), d_t is cash (M\$) and w_t is the total cash value (M€). Contracts made are shown by month $0, 1, 2, 3, 6$ and contracts maturing are shown by month $1, 2, 3, 6, 12$. Assets are shown initially and at the end of months $1, 2, 3, 6, 12$

Case	Month →	0	1	2	3	6	12	1	2	3	6	12
	Variable ↓	Contract and asset level						Contract maturing				
EqC												
0	$E(w_t)$	10.8	10.7	11.5	13.1	17.6	27.4					
1	$E(s_{ts})$	4.5	0.0	-0.0	1.1	1.6		1.2	1.1	2.2	1.1	1.6
	$E(p_{xts})$	0.0	0.0	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0
	$E(e_t)$	10.0	10.8	11.5	13.1	17.7	27.5					
	$E(w_t)$	10.8	10.8	11.5	13.1	17.7	27.5					
2	$E(s_{ts})$	10.0	0.2	0.6	1.7	1.9		0.0	0.0	0.0	5.6	8.8
	$E(b_{ts})$	0.0	0.7	2.1	3.8	2.7		0.0	0.0	0.0	3.2	6.1
	$E(p_{xts})$	13.6	0.0	0.4	2.0	3.0		2.3	2.3	4.5	3.3	6.6
	$E(c_{xts})$	0.0	1.6	4.1	6.4	4.9		0.0	0.5	3.7	4.1	8.8
	$E(e_t)$	8.9	8.6	8.6	10.1	14.8	28.4					
	$E(d_t)$	0.0	1.4	2.4	2.8	2.8	0.0					
	$E(w_t)$	9.9	9.9	10.8	12.7	17.5	28.4					
TrE												
0	$E(w_t)$	10.8	10.8	11.6	13.2	17.6	26.8					
1	$E(s_{ts})$	1.2	2.7	0.7	2.6	2.7		1.2	1.1	2.2	2.6	2.7
	$E(p_{xts})$	0.0	0.0	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0
	$E(e_t)$	10.0	10.8	11.6	13.2	17.6	26.9					
	$E(w_t)$	10.8	10.8	11.6	13.2	17.6	26.9					
2	$E(s_{ts})$	5.0	6.4	3.8	3.9	2.3		0.0	0.0	0.0	5.3	16.1
	$E(b_{ts})$	0.0	0.0	0.1	2.1	1.9		0.0	0.0	0.0	1.3	2.8
	$E(p_{xts})$	5.4	2.3	1.6	2.5	2.0		0.0	2.3	4.4	3.0	3.9
	$E(c_{xts})$	0.0	0.5	1.4	3.6	3.6		0.0	0.5	1.3	2.5	4.9
	$E(e_t)$	9.6	9.3	9.4	11.1	15.5	27.5					
	$E(d_t)$	0.0	1.0	1.9	1.9	2.1	0.0					
	$E(w_t)$	10.4	10.3	11.1	13.0	17.5	27.5					

References

W. Branson, Asset markets and relative prices in exchange rate determination. Sozialwissenschaftliche Annalen **1**, 69–89 (1977)

Y.-W. Cheung, M.D. Chinn, A.G. Pascual, Empirical exchange rate models of the nineties: are any fit to survive? J. Int. Money Financ. **24**(7), 1150–1175 (2005)

R. Dornbusch, Expectations and exchange rate dynamics. J. Polit. Econ. **84**, 1161–1176 (1976)

C. Engel, K.D. West, Exchange rates and fundamentals J. Polit. Econ. **113**(3), 485–517 (2005)

C.M. Engel, N.C. Mark, K.D. West, *Exchange Rate Models Are Not as Bad as You Think* (National Bureau of Economic Research, Cambridge, 2007)

C. Engel, N.C. Mark, K.D. West, Factor model forecasts of exchange rates. Econ. Rev. **34**(1-2), 32–55 (2015)

R. Engle, C. Granger, Co-integration and error correction: representation, estimation and testing. Econometrica **55**(2), 251–276 (1987)

J.A. Frenkel, A monetary approach to the exchange rate: doctrinal aspects and empirical evidence. Scand. J. Econ. **76**(May), 200–224 (1976)

J.K. Galimberti, M.L. Moura, Taylor rules and exchange rate predictability in emerging economies. J. Int. Money Financ. **32**, 1008–1031 (2013)

J. Hull, *Options, Forwards and Other Derivatives*, 5th edn. (Prentice Hall, Upper Saddle River, 2002)

O. Ince, Forecasting exchange rates out-of-sample with panel methods and real-time data. J. Int. Money Financ. **43**, 1–18 (2014)

K.K. Lewis, Testing the portfolio balance model: a multi-lateral approach. J. Int. Econ. **24**(1), 109–127 (1988)

R. MacDonald, M.P. Taylor, The monetary model of the exchange rate: long-run relationships, short-run dynamics and how to beat a random walk. J. Int. Money Financ. **13**(3), 276–290 (1994)

N.C. Mark, Exchange rates and fundamentals: evidence on long-horizon predictability. Am. Econ. Rev. **85**(1), 201–218 (1995)

N.C. Mark, Changing monetary policy rules, learning, and real exchange rate dynamics. J. Money Credit Bank. **41**(6), 1047–1070 (2009)

N.C. Mark, D. Sul, When are pooled panel-data regression forecasts of exchange rates more accurate than the time-series regression forecasts?, in *Handbook of Exchange Rates*, ed. by J. James, I.W. Marsh, L. Sarno (John Wiley & Sons, 2011), pp. 265–281

R.A. Meese, K. Rogoff, Empirical exchange rate models of the seventies: do they fit out of sample? J. Int. Econ. **14**(1), 3–24 (1983a)

R.A. Meese, K. Rogoff, The out-of-sample failure of empirical exchange rate models: sampling error or misspecification?, in *Exchange Rates and International Macroeconomics*, ed. by J.A. Frenkel (University of Chicago Press, Chicago, 1983b), pp. 67–112

T. Molodtsova, D.H. Papell, Out-of-sample exchange rate predictability with Taylor rule fundamentals. J. Int. Econ. **77**(2), 167–180 (2009)

T. Molodtsova, A. Nikolsko-Rzhevskyy, D.H. Papell, Taylor rules and the euro. J. Money Credit Bank. **43**(2-3), 535–552 (2011)

R.J. Sweeney, Mean reversion in G-10 nominal exchange rates. J. Financ. Quant. Anal. **41**(3), 685–708 (2006)

J.B. Taylor, Discretion versus policy rules in practice. Paper presented at the Carnegie-Rochester conference series on public policy (1993)

W.T. Ziemba, J.M. Mulvey (eds.), *Worldwide Asset and Liability Modeling* (Cambridge University Press, Cambridge, 1998)